DEVELOPMENT OF COMBINATION THERAPY FOR PROSTATE CANCER

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alexander R. Root

July 2017

DEVELOPMENT OF COMBINATION THERAPY FOR PROSTATE CANCER

Alexander R. Root, Ph.D.

Cornell University 2017

There is an unmet need in prostate cancer for effective therapies to prevent the emergence of resistance. Combinations of small molecules targeting key pathways are a promising strategy. I investigated how populations of the early metastatic prostate cancer cell line LNCaP respond at the proteomic and phenotypic levels to six clinically-relevant, one-drug treatments and their 15 pairs of two-drug combinations, administered simultaneously to treatment-naive cells. After 24 hours of drug addition for all 21 drug treatments I measured 52 total-proteins by selected-reaction monitoring mass-spectrometry based proteomics (SRM), 20 phospho-proteins, and 50 total-proteins by reverse-phase protein arrays (RPPA). I measured phenotypic effects on cell proliferation and apoptosis in all conditions using phase-contrast and fluorescence microscopy. Network analysis identified (phospho)-proteins with large responses to drug treatments that are druggable with FDA-approved drugs or have nearest-neighbors that are druggable.

A total of ten drugs targeting these nearest responder (phospho)-proteins were tested in single, double, and triple combinations. I found that 7 out of 10 triple combinations co-targeting androgen receptor and PI3K pathways were no more effective than the two-drug combination at the doses tested: PRKC (enzastaurin), MAPK (losmapimod), STAT3 (napabucasin), HDAC (panobinostat), SRC (saracatinib), casein kinase (silmitasertib), MAPK (ulixertinib). I was unable to determine the relative efficacy of aurora kinase B (barasertib), 14-3-3 in-

hibor (BV02) and their combinations. For one drug tested (dinaciclib, CDK1/2 inhibitor) I found increased efficacy with PI3K signaling inhibition (MK2206, AKT1/2 inhibitor).

I propose a simple model where blocking PI3K signaling in LNCaP cells causes a partial inhibition of the cell cycle G1/S transition through hypo-phosphorylation of RB and hypo-phosphorylation of CDK2 at the G2/M transition. These effects can be enhanced by co-targeting PI3K signaling and drivers of the cell cycle using the CDK1/2 inhibitor dinaciclib.

I also investigated how proteins that interact with the androgen receptor change in response to perturbation in a prostate cancer context (VCaP cell line). I identified 111 proteins with affinity purification mass spectrometry (AP-MS). These proteins have diverse functions and many have not been reported in the literature. I found that treatment with an antagonist (enzalutamide) has minimal effects on the interactors. Interactors such as metadherin (MDTH) and chromosome transmission fidelity factor 8 (CHTF8) are recurrently altered in patient cohorts. Metadherin is recurrently amplified and may represent a promising drug target.

BIOGRAPHICAL SKETCH

My undergraduate major was in biomedical engineering, which included an additional specialization, sub-major in tissue engineering. As an undergraduate, I performed research in the laboratory of Eduardo Macagno on mechanisms of neuronal growth and arbor formation using a quantitative approach. During the 2nd Iraq and Afghanistan wars I pursued public service as a mathematics teacher and worked in public schools where all students came from households below the federal poverty line.

My thesis research in the laboratory of Chris Sander involved preclinical systems pharmacology and systems biology. I worked on the development of combination therapy for early stage metastatic prostate cancer using a combined wet- and dry-lab approach. My thesis research included training and practice in experimental methods and mathematical modeling. I performed the majority of work in steps 1-5 with the exception of the targeted proteomics, which was performed by collaboration with Ruedi Aebersold's laboratory. In addition to my graduate coursework, I received training in cell culture and pharmacology assays from members of the Sander laboratory and the Molecular Cytology and Microscopy core facility; training in targeted proteomics from a rotation in the laboratory of Paul Tempst and several visits to the laboratory of Ruedi Aebersold; and training in network modeling from members of the Sander laboratory.

I am interested in continuing research on the development of combination therapy in post-doctoral work along several lines: (1) use of newer model systems such prostate organoids; (2) use of additional model systems to assess drug combination toxicity, such as primary human hepatocytes; (3) better selection and measurement of protein targets through methods development; (4) measuring drug response over multiple time points; (5) extending network model-

ing approaches to incorporate this temporal data; (6) greater use of automation technology to improve sample throughput. I anticipate the greatest challenges and unmet needs are (3) better selection and measurement of protein targets through methods development in targeted proteomics.

Dedicated to Paul Dyckman for his inspiring teaching.

ACKNOWLEDGEMENTS

Chris Sander for mentorship, training, and inspiration. H. Alexander Ebhardt for training. Members of the Sander and Aebersold laboratories for helpful questions and comments. Thesis committee members including Mark Rubin and Paul Tempst for training and constructive criticism. Physiology, Biophysics, and Systems Biology department members for education. Laboratory rotation supervisors Chris Sander, Paul Tempst, Joan Massague, Francesca Demichelis, and Olivier Elemento for training and mentorship.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . v
Acknowledgements . . . vi
Table of Contents . . . vii
List of Abbreviations . . . ix
List of Symbols . . . ix
List of Tables . . . x
List of Figures . . . xii

1 Introduction: Designing Combination Therapy for Prostate Cancer 1
1.1 Why combination therapy? . . . 1
1.2 The perturbation biology approach . . . 2
1.2.1 Hypothesis . . . 4
1.2.2 Aims . . . 4

2 Methods - Design of Perturbations, Measurements and Analysis 7
2.1 Perturbations . . . 7
2.1.1 LNCaP cell line model . . . 7
2.1.2 Drug selection and treatments . . . 8
2.2 Measurement methods: phenotypic and proteomics quantitation of drug responses . . . 10
2.2.1 Microscopy methods for phenotypic responses to drugs . 10
2.2.2 Targeted proteomics methods for drug effects on proteins and phosphosites . . . 11
2.3 Network analysis methods, predictions, and assessment . . . 18
2.3.1 Nominating novel drug combinations with network analysis . . . 18
2.3.2 Methods for testing network analysis predictions with phenotypic measurements of cell death . . . 19
2.3.3 Methods for testing network analysis predictions with molecular measurements by discovery proteomics . . . 21

3 Results - Quantifying Molecular and Phenotypic Responses to Drug Perturbations 24
3.1 Establishment of SRM assays . . . 24
3.2 Phenotypic and proteomic responses to perturbations . . . 26
3.3 Network analysis and predictions . . . 26

4 Results & Discussion - Strategy 1: Co-targeting PI3K Signaling and Cell Cycle 30
4.1 Testing predictions uncovers a PI3K signaling & cell cycle inhibition strategy . . . 30
4.2 Discussion on PI3K signaling & cell cycle inhibition strategy . . . 30

5 Results & Discussion - Strategy 2: Co-targeting 14-3-3 Proteins in Combination with AKT and AR **34**
5.1 Testing predictions uncovers a 14-3-3 PI3K, and AR inhibition strategy . . . 34
5.2 Discussion on 14-3-3, PI3K, and AR inhibition strategy . . . 35

6 Alternative Approach: Identifying Androgen Receptor Interacting Proteins **43**
6.1 Introduction - AR-interacting proteins may be good drug targets 43
6.2 Methods - identification of AR-interacting proteins in a cellular context relevant to prostate cancer with AP-MS . . . 45
6.2.1 Cell culture and drug treatments . . . 45
6.2.2 Affinity Purification Mass Spectrometry . . . 47
6.3 Results - identification of AR interactors and effects of drug perturbations . . . 50
6.4 Discussion . . . 55

7 Conclusion & Future Directions for Prostate Cancer Combination Therapy **59**
7.1 Targeted proteomics and perturbation biology are a promising approach to develop combination therapy . . . 59
7.2 A diversity of approaches to develop combination therapy should continue to be pursued . . . 61

A Appendix - Review Article Reprint: Applications of Targeted Proteomics in Systems Biology and Translational Medicine **64**

B Appendix - Note on collaborations and authorship **81**

C Appendix - Mathematical modeling of drug effects **82**
C.1 Partial least-squares regression model . . . 83
C.1.1 Introduction & Methods . . . 83
C.1.2 Results . . . 85
C.2 Undirected network model with partial correlations . . . 87
C.3 Coupled-system of nonlinear ordinary differential equations . . . 88

Bibliography **98**

LIST OF ABBREVIATIONS

Instruments & techniques

AP	affinity purification
MS	mass spectrometry or mass spectrometer
AP-MS	affinity-purification mass spectrometry
GEM	genetically engineered mice
MS/MS	tandem mass spectrometry
SRM	selected reaction monitoring
RPPA	reverse-phase protein array
SVD	singular value decomposition
PLSR	partial least-squares regression
BP	belief propagation

First perturbation drug set

DAS	dasatinib, SRC inhibitor
DMSO	dimethyl sulfoxide, drug vehicle
DOC	docetaxel, blocks microtubule depolymerization
ENZ	enzalutamide, AR inhibitor
MDV	MDV3100, AR inhibitor
MK2	MK2206, AKT1/2 inhibitor
TEM	temsirolimus, MTOR inhibitor

Second perturbation drug set

BAR	barasertib, aurora kinase inhibitor
DIN	dinaciclib, CDK1/2/5 inhibitor
STA	enzastaurin, PRKDC inhibitor
LOS	losmapimod, MAPK inhibitor
NAP	napabucasin, STAT3 inhibitor
PAN	panobinostat, HDAC inhibitor
SAR	saracatinib, SRC inhibitor
SIL	silmitasertib, casein kinase inhibitor
ULI	ulixertinib, ERK1/2 inhibitor

AR interactors perturbation set

DHT	dihydrotestosterone, AR ligand

LIST OF SYMBOLS

$\mathbf{W}$	matrix of interactions or connectivity between proteins
w_{ij}	each entry in the interaction matrix $\mathbf{W}$
w_{ij}	each entry in the interaction matrix $\mathbf{W}$

LIST OF TABLES

2.1 Top-ranked predictions from druggable, nearest-responder-node analysis. References are given for the target or the drug. Doses are reported or estimated IC50 values from phenotypic assays. 19

LIST OF FIGURES

1.1 **Overview of the experimental design, measurements, network analysis predictions, and testing of analysis predictions. (A)** To develop combination therapy for AR-positive prostate cancer I perturbed LNCaP cells with six-clinically relevant drugs and their corresponding 15 drug pairs. **(B)** Measurements are short-term drug response of populations of cells with targeted proteomics and phenotypic assays. **(C)** Data are placed on a human protein-protein, protein-drug network from the PathwayCommons database [Cerami et al., 2011]. Network analysis nominated new drug combinations that may be effective alone or in combination with the original drug set tested. **(D)** Testing of network predictions lead to promising results for 1 out of 9 drugs tested. 6

2.1 **Drug-target selection and doses.** Drugs were chosen by manual curation of the literature and clinical trials for clinically-relevant drugs. Several drugs were chosen targeting PI3K/AKT signaling because PTEN loss is a recurrent alteration in prostate cancer and in the LNCaP cell line. Drugs had to have passed phase I trials in the USA for any indication. Although docetaxel is a chemotherapeutic agent and not a targeted agent, it was chosen because it is a standard of care. Doses were chosen based on manual literature curation of IC50s in LNCaP or other cell lines if data were not available for LNCaP. 9

2.2 **Origins of candidate proteins for SRM assay development.** . . 12

3.1 **SRM assay development and exploratory perturbation for time point selection.** *Left panel*. Violin plot showing protein abundance distribution of proteins quantified in LNCaP cells from Geiger et al., this study and estimated abundance values of the target PLIPCa list. Circular Interaction Graph for Proteomics representing the origin of all peptides corresponding to the target PLIPCa proteins. *Middle panel*. SRM assay development. Top row showing sum of transitions per peptide for three proteins (AR, KLK3 and CYC). Endogenous peptides co-elute with stable isotope labelled reference peptides. Bottom row showing individual transitions per peptide which are base line separated. *Right panel*. Time point selection. Top: six drugs were chosen and automated fluorescent microscopy quantified for apoptosis as a function of time. Different drugs appear to trigger apoptosis with varying kinetics. Therefore, our choice of the 24 hour timepoint is arbitrary and misses the peak signal for many conditions. Bottom: the proteome of LNCaP cells treated with enzalutamide was quantified at 0.5, 6 and 24 h post treatment. The relative abundance changes for four proteins are shown. 25

3.2 **Phenotypic and proteomic responses to one- and two-drug treatments. (A)** Phenotypic effects of one- and two-drug combinations administered concurrently at single doses were assessed by microscopy. Drug names, targets and doses are shown in Table 1. Drug effects were quantified by percent area of viable cells, which is computed by subtracting the area of fluorescence channel signal of the apoptosis probe (caspase 3/7 activation) from the phase contrast channel measurement of cell area. Cells grew in clumps that could not be segmented into individual cells for a computation of percent apoptotic cells. Comparing drug effects at day 5 (right panel) to day 1 (left panel) reveals that effects are more pronounced and more varied at day 5 than day 1. Error bars are +/- SEM. **(B)** Integrated log2 responses of targets compared to vehicle 24 hours after drug treatment. Proteins were measured by selected-reaction monitoring mass spectrometry (SRM) and additional (phospho)-proteins were measured by reverse-phase protein arrays (RPPA). 27

3.3 **Network analysis of drug treatment responses identifies potential network vulnerabilities and new drug combination strategies.** **(A)** The analysis method has three steps: (1) drug response data are mapped onto a database network of protein-protein and protein-drug interactions from PathwayCommons.org; (2) nearest-neighbor subnetworks around proteins with strong responses to perturbations are extracted for visualization and further analysis; (3) an additional filter is applied for commercial availability and the extent of literature on the role of the target and/or efficacy of the drug in prostate cancer because there are typically numerous options for drugging the nearest node to a (phospho)-protein with strong responses. **(B)** Phospho-RB has strong responses to numerous perturbations including AR inhibition with enzalutamide. This subnetwork shows that there are no available drugs for targeting pRB directly but several options for drugging its nearest neighbors. 29

4.1 **Testing network analysis predictions of the efficacy of CDK inhibition in a cell death assay.** The CDK1/2/5 inhibitor dinaciclib was tested in all one-, two-, and three-drug combinations coadministered at single doses with 4 members of the first drug set: docetaxel (microtuble inhibitor), enzalutamide (AR inhibitor), MK2206 (AKT inhibitor), and temsirolimus. Cell death was assessed after 48 hours of drug treatments. Doses are given in Table 2. Dinaciclib is most effective in combination with MK2206 and MK2206 with temsirolimus. Dinaciclib is relatively effective on its own. Error bars are +/- SEM. Nine additional drug predictions from network analysis were assayed similarly in all one-, two-, and three-drug combinations but seven did not show efficacy at the dose tested and for two drugs effects could not be determined. 31

4.2 **Combined inhibition of AKT and cell cycle is more effective than either drug alone.** Following single drug treatment with PI3K pathway inhibitors network analysis of targeted proteomics measurements of drug responses led to the hypothesis that augmenting the response of phospho-RB by inhibition of CDK1/2 would enhance drug efficacy. Single drug efficacy may be limited by a feedback loop operating between MTOR and PI3K or between AR and CDK4. A drug targeting CDK1/2 (dinaciclib DIN) was chosen based upon commercial availability, status in clinical trials, and potency. A cell death assay confirmed the prediction. This inhibition strategy is similar to serial pathway inhibition, however, the effects are probably more complicated because of the many interactions among these and other cell cycle proteins not shown. 33

5.1 **Patterns of protein responses across perturbations identify consistent strong responders. I**. Proteins frequently quantified in prostate cancer (e.g. KLK3 [PSA] or HSP90) show relatively small changes across pharmacologically perturbed conditions. **II**. Proteins consistently decreasing as a function of perturbation. PAK1 and HN1 have not been shown to be important for prostate cancer. **III**. Proteins with consistent increases following drug perturbation include 14-3-3 proteins. 36

5.2 **Recurrent alterations in primary and metastatic tumors show amplifications in 14-3-3 protein YWHAZ. (A)** Oncoprint of primary prostate tumor data shows recurrent alterations in PTEN and FOXO1. **(B)** Oncoprint of metastatic prostate tumor patient data shows recurrent amplifications in AR and YWHAZ. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]] . 39

5.3 **Effects on relative protein expression of 14-3-3 inhibitor BV02 for proteins in the target list.** 40

5.4 **Geneset analysis of triple combination therapy of ARi, AKTi, and YWHAZi indicates decreases in DNA repair and other processes.** 41

5.5 **Combination therapy targeting YWHAZ, AKT and AR and patient alterations in YWHAZ. (A)** Molecular and phenotypic effects of YWHAZ inhibitor BV02 **(B)** YWHAZ amplifications cooccur with MYC on 8q or 2p. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]] 42

6.1 **Identification of AR-interacting proteins after ligand, agonist, and antagonist treatment by affinity purification mass spectrometry (AP-MS). (A)** The prostate cancer cell line VCaP is an early metastatic model that expresses full-length AR and is sensitive to its inhibition. VCaP cells were grown in normal culture media and treated with ligand dihydroxytestosterone (DHT), agonist bicalutamide (BICA), and antagonist enzalutamide (ENZA). Cells were lysed and fractionated into cytoplasmic- and nuclear-enriched. A polyclonal antibody recognizing the N-terminal domain of AR immunoprecipitated interacting proteins. Analysis by mass spectrometry revealed a number of proteins differentially abundant across treatments. Additional analyses integrating databases of AR-interacting proteins, recurrent alterations in prostate cancer patients, and druggability prioritized interactors for additional experimentation. **(B)** Domain structure of full-length androgen receptor including the N-terminal domain (NTD), DNA-binding domain (DBD), hinge, and ligand-binding domain (LBD). The LBD is thought to be the site of drug binding. The antibody used for pull-down is Santa Cruz N20 binds in the first 50 amino acids potentially disrupting some protein-protein interactions. Other methods, such as FLAG- or HIS-tags would also suffer from this problem. . . . 46

6.2 **AR antibody pulls down complexes and negative control IgG pull down does not. (A)** Western blot for AR following drug vehicle (V), ligand DHT (D), agonist bicalutamide (B), and antagonist enzalutamide (E) for nuclear- and cytoplasma-enriched fractions. **(B)** ImmunoPrep of AR in the experiment shows reasonable efficiency using the N20 antibody. Control IgG is not pulling down AR. 49

6.3 **Silver stain of AR pull-downs following perturbations shows numerous interactors but few changes across perturbations.** There are a few observable changes in bands across conditions but in general not large differences among conditions, with the caveat that silver stain is not always a particularly sensitive technique. (V) drug vehicle, (D) agonist DHT, (B) bicalutamide, (M) antagonist MDV3100 or enzalutamide. The negative control lane shows few bands as expected. 50

6.4 **AP-MS identifies numerous, diverse, and novel AR-interacting proteins. (A)** Log2 of protein spectral counts for proteins with a Bayesian false-discovery-rate less than 0.05 compared to IgG negative control as determined by SAINTexpress. Missing values shown in gray do not necessarily indicate the absence of a proteins expression. **(B)** Simplified ontology categories for the AP-MS proteins show a large diversity in functions. GO categories for all the proteins were manually simplified into eight categories. **(C)** Overlap of AP-MS with known interactors reveals novel interactors. AR interactors reported in several databases show limited agreement with interactors found by AP-MS in VCaP. PathwayCommons (PC8) is a union of public pathway databases; String (version 10) was limited to high confidence, experimentally determined interactors; McGill is a manually curated database of AR interactors. **(D)** Computationally predicted probability of AP-MS interactions determined by PrePPI method, which uses structural, functional, evolutionary, and protein expression (Chen/Honig, PLoS Comp Bio 2015) . . 52

6.5 **Statistical assessment of differential expression between negative control vs. perturbations.** Peptide-to-protein summarization and quantification of spectral counts is performed with the Transproteomic Pipeline search engines. Statistical assessment comparing each perturbation condition against IgG pulldown was performed with SAINTexpress. It computes a Bayesian false-discovery-rate (BFDR) for each protein, corrected for multiple hypothesis testing. Proteins with high spectral counts are more likely to show statistically significant differences with negative control IgG pull-down. 53

6.6 **Occurrences of AR-interaction motifs in interactors. Androgen receptor is known to bind to LXXLL and FXXLF motifs.** Approximately 1/3 of the interactors identified by AP-MS contain at least 1 of these motifs. PRKDC is an outlier containing numerous occurrences of the LXXLL motif. Presence of either motif is evidence that an identified interactor is less likely to be a false positive. For example, BAG2 and CAM2KD are not reported as AR interactors in PathwayCommons, McGill AR database, or StringDB but they each contain an LXXLL motif. 54

6.7 **Distribution and overlap of AR interaction motifs among identified interactors.** 55

6.8 **AR-Interacting proteins are present in similar abundances across ligand and antagonist treatments. (A)** Plot of ligand DHT vs antagonist enzalutamide shows most proteins have similar expression levels in both treatments. Statistical analysis with SAINTexpress indicated no statistically significant differences in expressed proteins across the treatment conditions. SAINTexpress incorporates spectral counts and percent of a proteins amino acid chain detected. **(B)** Six proteins are shown for biological interest. Black dots indicate measurements in 4 biological replicates. Blue dots indicate mean values and blue lines indicate +/- 1.5 standard deviations. BAG2 promotes protein release from the chaperone HSP70. CTNNA1 works in cell adhesion by interacting with cadherins. GMPPB is thought to be involved in oligosaccharide metabolism. PRKDC is a serine/threonine kinase that senses DNA damage. SMARCE1 is part of the SWI/SNF complex and is involved in chromatin modification. USP7 is a deubitquitinase. 56

6.9 **Recurrent alterations in AR-interacting proteins in large clinical cohorts.** Most AR interacting proteins are recurrently altered in less than 20% of patient samples in primary and metastatic prostate cancer at the levels of DNA mutations, copy number deletions/amplifications and mRNA expression high/low. Data for primary prostate tumors comes from a TCGA 2015 study with more than 300 patient samples. Metastatic tumor data in Robinson et al., Cell 2015 (Michigan) has more than 100 patient samples. In metastatic tumors amplifications occur in metadherin MTDH, which is known to activate NF-kappaB. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]] 57

C.1 **PLSR performance on withheld data indicates poor predictive power for most drug combinations except combinations including temsirolimus.** 86

C.2 **Network of partial correlations showing top 50 strongest edges.** 89

C.3 **Workflow for Hopfield network modeling with belief propagation algorithm, FORTRAN code, and R code.** The flow starts in the upper left corner and moves diagonally to lower right. . . 93

C.4 **Tuning of the sparsity parameter by assessing performance on withheld data for the top 60 models with lowest training error.** 94

C.5 **Performance of top ten BP-models with lowest training error on withheld data by Pearson correlation data vs model.** 95

C.6 ***In silico* perturbation predictions on cellular phenotypes and KLK3 show no good two-drug combinations to effect cell phenotypes but possibly strategies to minimize KLK3.** 96

C.7 ***In silico* perturbation predictions on KLK3 minimization reveals multiple two-drug combination strategies.** 97

CHAPTER 1

INTRODUCTION: DESIGNING COMBINATION THERAPY FOR PROSTATE CANCER

1.1 Why combination therapy?

There is an unmet need in prostate cancer for effective therapies against metastatic disease [Attard et al., 2016]. Recently, studies of large patient cohorts have identified numerous recurrent alterations [Taylor et al., 2010]. These alterations form the basis for current targeted therapies against AR signaling and a rationale for novel therapy development against cell cycle, DNA repair, ETS fusions, chromatin remodelling, MAPK pathway, Wnt pathway, and PI3K pathway [Spratt et al., 2016], [Rodrigues et al., 2017]. The emergence of resistance to current therapies targeting androgen receptor signaling and taxane chemotherapy may be overcome with combination therapy [Watson et al., 2015], [Komura et al., 2016], [Dumontet and Sikic, 1999]. The ability to overcome the emergence of resistance with combination therapy treating HIV and tuberculosis suggests it may be a promising approach for prostate cancer [Bock and Lengauer, 2012]. Theoretical evolutionary dynamics studies also suggest combination therapy may be able to overcome the emergence of resistance [Bozic et al., 2013]. Drug combinations that are effective against tumor cells may not cause a concomitant multiplicative increase in toxicity [Lehár et al., 2009].

Development of combination therapy is challenging partly due to a search space that grows exponentially with the library size, number of drugs in a combination, doses, exposure history, administration schedule (i.e.,

whether the drugs are given simultaneously or staggered), and other factors [Zimmer et al., 2016]. One pre-clinical approach to developing combination therapy, termed perturbation biology investigates how populations of treatment-naive cells in monoculture respond to one- and two-drug treatments at the level of (phospho)-proteins and cellular phenotypes after short-term exposure; these short-term (phospho)-proteomic responses may represent resistance mechanisms that may be targeted in novel drug combinations leading to increased efficacy [Korkut et al., 2015]. Theoretical evolutionary dynamics has identified tumor population and growth rate as key factors in the probability that resistance emerges; therefore, drug combinations that decrease tumor population and/or growth rate may prevent the emergence of resistance [Bozic et al., 2013].

1.2 The perturbation biology approach

The perturbation biology approach is a promising strategy to develop combination therapy and learn how cellular networks respond to and resist small molecular perturbations [Korkut et al., 2015]. Perturbation biology involves measuring how prostate cancer cells respond to clinically-relevant drug perturbations to nominate mechanisms of resistance that in turn may be overcome with drug combinations – an approach that has worked previously in BRAF-resistant melanoma [Korkut et al., 2015].

The most useful molecular measurements to uncover short-term resistance mechanisms may be (phospho)-proteins because proteins are the targets of most drugs and are key components of biochemical networks

[Aebersold and Mann, 2016]. Measurements of (phospho)-proteins after drug response led to a key discovery in the understanding of prostate cancer signal transduction networks, where reciprocal feedback inhibition between AR and PI3K signaling provides resistance mechanisms to one-drug treatments targeting either pathway [Carver et al., 2011]. Additionally, proteomics has provided numerous insights in cancer including subtyping [Tyanova et al., 2016], connecting somatic mutations to signaling [Mertins et al., 2016], and understanding the functions of genes [Wang et al., 2017]. Of the various types of proteomics, targeted proteomics provides favorable sample throughput and sensitivity for perturbation biology studies [Ebhardt et al., 2015], [Lu et al., 2016].

To develop a combination therapy for prostate cancer I chose the LNCaP cell line model of early metastatic prostate cancer; it is PTEN-deficient, AR sensitive, and has an ETS-family rearrangement; all of which are recurrent alterations in early metastatic prostate cancer patients [Taylor et al., 2010]. LNCaP was derived from the lymph node of a patient with metastatic prostate cancer and is sometimes considered the dominant model system in the field [Watson et al., 2015]. I chose to use two complementary platforms for targeted proteomics: selected reaction monitoring mass spectrometry (SRM) and reverse-phase protein arrays (RPPA). I performed manual literature curation to choose targets and developed SRM assays for 52 total-protein measurements. To increase coverage of cancer-relevant signaling proteins and phosphosites I used a commonly used set of RPPA antibodies [Lu et al., 2016]. I selected six clinically-relevant drugs and their doses based on a literature search: AR inhibitor enzalutamide [Tran et al., 2009], chemotherapy docetaxel [Dumontet and Sikic, 1999], SRC inhibitor dasatinib [Nam, 2005], AKT inhibitor MK2206 [Yap et al., 2011], and MTOR inhibitor temsirolimus

[Wang et al., 2011].

After performing all one- and two-drug combinations I used network analysis of the (phospho)-proteins responses to nominate resistance mechanisms and additional drug combinations to overcome them. One limitation to this approach of studying LNCaP cells in monoculture is that microenvironment cells have been shown to secrete factors that alter drug sensitivity and provide resistance mechanisms [Ireland et al., 2016].

1.2.1 Hypothesis

Measuring how LNCaP cells respond to short-term one- and two-drug combinations at the phenotypic and proteomic levels will reveal markers of response and resistance that can be enhanced or blocked with novel drug combinations leading to a highly potent cocktail. Such a potent cocktail works by lowering the probability of resistance emerging through: shrinking the tumor population size and/or targeting pre-existing resistant clones. A cocktail composed of targeted therapies will have much greater effects on the tumor than concomitant increases in toxicity to normal cells.

1.2.2 Aims

The first three aims were presented during my admission to candidacy exam (ACE). Additional aims arose during progression of the research.

1. Perform perturbation biology experiments using targeted proteomics:

- manually curate literature for proteins relevant for prostate cancer;
- manually curate literature for clinically-relevant drugs for prostate cancer;
- collaborate with the Aebersold laboratory to develop targeted proteomics assays (SRM);
- measure the phenotypic and proteomic effects of all one- and two-drug combinations.

2. Analyze the resulting data to nominate:
    - mechanisms of response and resistance;
    - novel drug combinations to enhance or overcome these mechanisms using network analysis.
3. Test network analysis predictions of novel drug combinations using phenotypic assays.
4. Evaluate novel drug combinations for testing *in vivo* mouse models.
5. Analyze perturbation data of AR-interacting proteins and prioritize candidates for drug combination development.

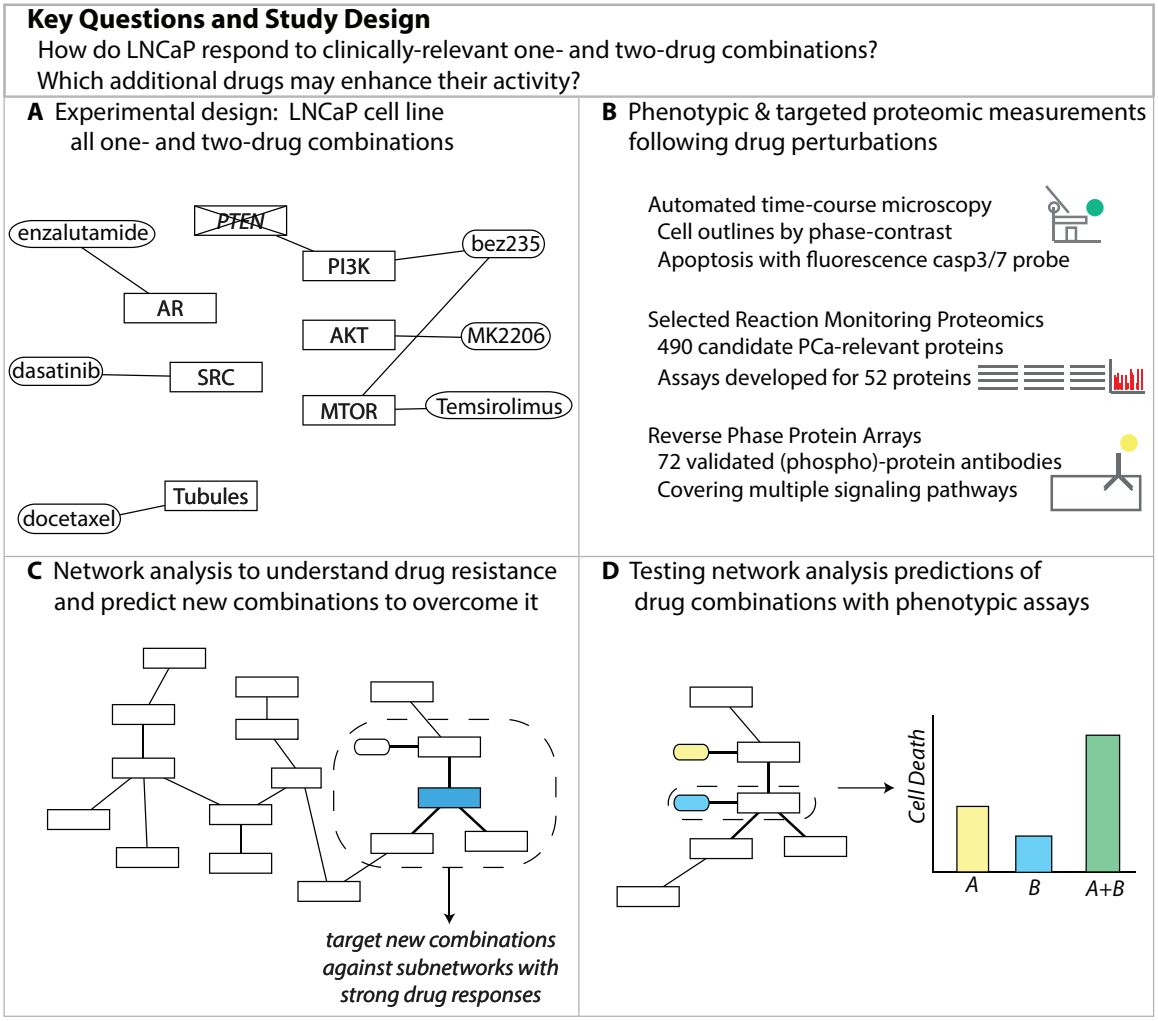

Figure 1.1: **Overview of the experimental design, measurements, network analysis predictions, and testing of analysis predictions.**
(A) To develop combination therapy for AR-positive prostate cancer I perturbed LNCaP cells with six-clinically relevant drugs and their corresponding 15 drug pairs.
(B) Measurements are short-term drug response of populations of cells with targeted proteomics and phenotypic assays.
(C) Data are placed on a human protein-protein, protein-drug network from the PathwayCommons database [Cerami et al., 2011]. Network analysis nominated new drug combinations that may be effective alone or in combination with the original drug set tested.
(D) Testing of network predictions lead to promising results for 1 out of 9 drugs tested.

CHAPTER 2

METHODS - DESIGN OF PERTURBATIONS, MEASUREMENTS AND ANALYSIS

2.1 Perturbations

An overview of the experimental design is depicted in figure 1.1. The process for developing targeted proteomics assays using SRM is described on the following pages in section 2.5.1 and figures 2.2 and 3.1.

2.1.1 LNCaP cell line model

The LNCaP cell line model of early metastatic prostate cancer was chosen for this research; it is PTEN-deficient, AR sensitive, and has an ETS-family rearrangement; all of which are recurrent alterations in early metastatic prostate cancer patients [Taylor et al., 2010]. LNCaP was derived from the lymph node of a patient with metastatic prostate cancer and is sometimes considered the dominant model system in the field [Watson et al., 2015].The prostate cancer cell line LNCaP clone FGC was ordered from ATCC and used throughout this study. Cultures were maintained in a humid 5% CO_2 atmosphere at 37°C and grown in RPMI-1640 medium supplemented with 10% heat-inactivated fetal bovine serum, 100 units/ml each of penicillin and streptomycin, and 2mM L-glutamine. Cells were used at less than 25 passages for all experiments. For phenotypic assays, cells were seeded in 96-well plastic bottom plates at 12,500 cells per well in biological triplicate for all conditions. For proteomics assays cells were seeded in 10cm petri dishes at 2 million cells per dish in biological

triplicate for all conditions. Cells were left at room temperature for 20 minutes for even seeding and placed in an incubator overnight.

2.1.2 Drug selection and treatments

Drug concentrations were estimated from the literature as IC50 values in phenotypic assays: bez235 500 nM, dasatinib 100 nM, docetaxel 10 nM, enzalutamide 10 µM, MK2206 1 µM, temsirolimus 100 nM. Drugs were used at the same concentration for singles and paired combinations. Drug dilutions were prepared such that the same volume of DMSO is added for each single, combination, and DMSO vehicle control. For proteomics samples, the 10cm dishes were placed in an incubator after drug addition and cells were scraped after 24 hours, pelleted, and frozen at -80°C. The sample layouts were not randomized and the experimenters were not blinded to the experimental conditions to minimize confusion during manual labeling and pipetting. For phenotypic measurements, a green-fluorescent, apoptosis-indicating reagent (NucView 488) was added concurrently with drugs to all wells. The 96-well plates were placed inside an automated microscope for live-cell imaging (IncuCyte ZOOM, Essen Biosciences), which resides in a humid 5% CO_2 atmosphere at 37°C incubator.

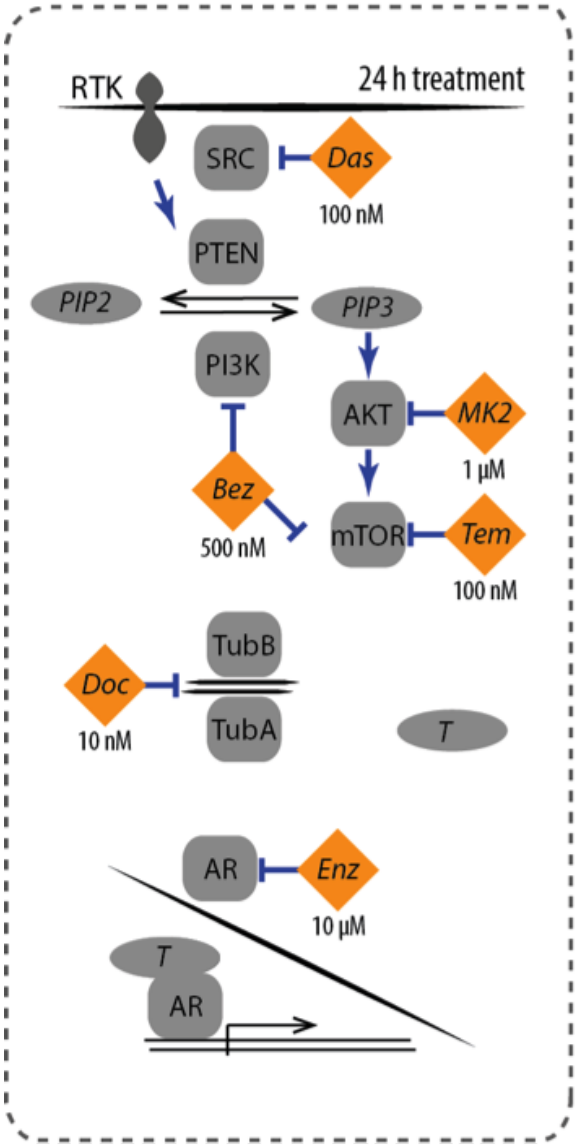

Figure 2.1: **Drug-target selection and doses.** Drugs were chosen by manual curation of the literature and clinical trials for clinically-relevant drugs. Several drugs were chosen targeting PI3K/AKT signaling because PTEN loss is a recurrent alteration in prostate cancer and in the LNCaP cell line. Drugs had to have passed phase I trials in the USA for any indication. Although docetaxel is a chemotherapeutic agent and not a targeted agent, it was chosen because it is a standard of care. Doses were chosen based on manual literature curation of IC50s in LNCaP or other cell lines if data were not available for LNCaP.

2.2 Measurement methods: phenotypic and proteomics quantitation of drug responses

2.2.1 Microscopy methods for phenotypic responses to drugs

Microscopy image acquisition and analysis

An automated, live-imaging microscope (IncuCyte ZOOM, Essen Biosciences) was set to record 4 images per well at 10x magnification every 4 hours following drug addition up to 120 hours using the control software (IncuCyte ZOOM v2015). Approximate outlines of cells were imaged using phase contrast and apoptosis was imaged using the green fluorescence channel. The microscope is automated to scan all wells of a 96-well plate and was set to default parameters for autofocus and exposure time. Images were analyzed using the same software. Several images were chosen to represent a broad spectrum of drug effects and controls to tune image analysis parameters, such that segmentation and thresholding appeared reasonable to the eye in both the phase contrast and green fluorescence channels. The image analysis pipeline was run in batch mode over all images in the time series [Chapman et al., 2016].

2.2.2 Targeted proteomics methods for drug effects on proteins and phosphosites

Label-free selected reaction monitoring assay development

note: SRM assay development was performed by H. Alexander Ebhardt in the Aebersold laboratory. I was invited by the Aebersold lab to oversee and assist in this development over the course of two weeks. I also learned aspects of SRM assay development during a laboratory rotation with Paul Tempst. I also co-authored a review article with H. Alexander Ebhardt, which is Appendix A.

Preparation of a list of candidate protein targets. To identify proteins relevant to prostate cancer I conducted a manual review of prostate cancer and general cancer literature. Proteins with evidence of differential mRNA abundance between primary metastatic samples were included. Approximately 400 proteins were mapped onto a human protein-protein interaction network and linker nodes were identified using the Netbox algorithm. In total 490 proteins were selected as candidates for targeted assay development shown in figure 2.2.

Sample preparation for label-free discovery MS/MS. Prostate cancer cells LNCaP (approximately 3 million) were resuspended in 800 µL of lysis buffer (8 M urea, 0.2% RapiGest, 0.1 M ammonium bicarbonate), vortexed for 10 s, and shaken at 1000 rpm at room temperature for 10 min. Samples were sonicated (amplitude 90%, cycle 0.6) for 2 x 2 min at 4°C and centrifuged at 13,200 rpm for 10 min. The clear lysate was transferred to a new tube, the amount of

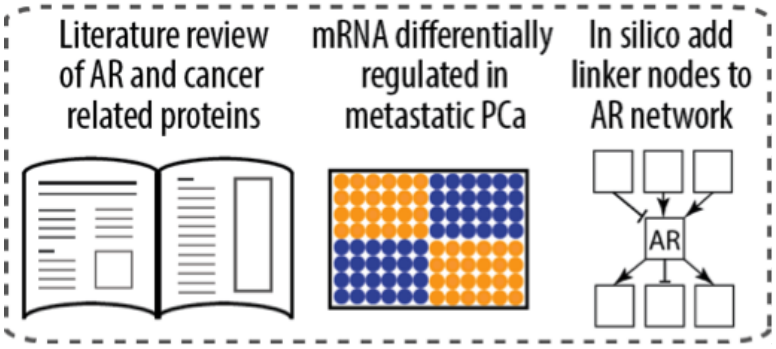

Figure 2.2: **Origins of candidate proteins for SRM assay development.**

protein was determined using the micro BCA assay (Pierce, Thermo Scientific), and the protein concentration was diluted to 5 mg/L with lysis buffer. The sample was reduced using TCEP (tris-2-carboxyethylphophine) and alkylated using iodoacetamide, and unreacted iodoacetamide was then neutralized with N-acetyl-cystein. The samples were then diluted with 0.1 M ammonium bicarbonate and 0.2% RapiGest to a final concentration of less than 2 M urea. Proteins were digested at a basic pH of 8.0 to 8.5 using sequence-grade modified trypsin (Promega) at a ratio of 1:100 for 14 h at 37°C. Digestion was stopped by decreasing the pH to less than 2 with formic acid. Samples were shaken another 30 min at 500 rpm at 37°C, followed by centrifugation at 13,200 rpm for 10 min to remove any insoluble fractions. Samples were cleaned with C18 columns.

Sample fractionation by strong cation exchange chromatography and isoelectric focusing. For strong cation-exchange fractionation, 500 µg of tryptic peptides were loaded on a PolyLC column (PolyWAX LP 200 x 2.1 mm, 5 µm, 300-A) for electrostatic repulsion liquid chromatography (ERLIC) fractionation using an Agilent 1200 HPLC chromatographic system. Buffer A: 90% ACN, 0.1% acetic acid, buffer B: 30% ACN, 0.1% formic acid. Flow rate: 0.5 mL/min. Gradient: 0-30% B in 40 min, 30-100% B in 40 min. One-minute fractions were

collected in a 96-well plate.

For isoelectric fractionation (OffGel, Agilent), 500 µg of tryptic peptides were loaded on a PolyLC column (PolyWAX LP 200 x 2.1 mm, 5 µm, 300-A) for electrostatic repulsion liquid chromatography (ERLIC) fractionation using an Agilent 1200 HPLC chromatographic system. Buffer A: 90% ACN, 0.1% acetic acid, buffer B: 30% ACN, 0.1% formic acid. Flow rate: 0.5 mL/min. Gradient: 0-30% B in 40 min, 30-100% B in 40 min. One-minute fractions were collected in a 96-well plate.

Label-free discovery MS/MS. Fractions were pooled into 10 samples for each fractionation method separately. 10% of each sample was analyzed by LC MS/MS in an Orbitrip LTQ Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to nano-LC (Proxeon, Odense, Denmark) equipped with a a reversed-phase chromatography 12-cm column with an inner diameter of 75 µm, packed with 5 µm C18 particles (Nikkyo Technos, Japan). Chromatographic gradients started at 97% buffer A and 3% buffer B with a flow rate of 300 nL/min and gradually increased to 93% buffer A and 7% bufferB in 1 min and to 65%bufferA and 35% buffer B in 60 minutes. After each analysis, the column was washed for 10 min with 10% buffer A and 90% buffer (buffer A: 0.1% formic acid in water; buffer B: 0.1% formic acid in acetonitrile). The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.2 kV and source temperature at 250 C. Ultramark 1621 for the FT mass analyzer was used for external calibration prior the analyses. Moreover, an internal calibration was also performed using the background poly-siloxane ion signal at m/z 445.1200.

The instrument was operated in DDA mode and full MS scans with 1 microscan at resolution of 60,000 were used over a mass range of m/z 2502000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1e6, dynamic exclusion (60 s with a repeat duration of 30 s [repeat count = 1], and an exclusion list size of 500 ions), and the charge state filter disqualifying singly charged peptides was activated. Following each survey scan, the top 20 most intense ions with multiple charged ions above a threshold ion count of 5000 were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision- induced dissociation (CID) were acquired in the linear ion trap, AGC was set to 5e4, isolation window of 2.0 m/z, and activation time of 0.1 ms, and maximum injection time of 100 ms was used. All data were acquired with Xcalibur software v2.2.

Acquired data were analyzed using the Proteome Discoverer software suite (v1.3.0.339, Thermo Fisher Scientific), and the Mascot search engine (v2.3, Matrix Science) was used for peptide identification. Data were searched against an in-house generated database containing all proteins corresponding to human in the Swissprot human database (as of 3/2013) with 600 added common contaminants, with a total number of 37,694 sequences in the database. A precursor ion mass tolerance of 7 ppm at the MS1 level was used, and up to three miscleavages for trypsin were allowed. The fragment ion mass tolerance was set to 0.5 Da. Oxidation of methionine and protein acetylation at the N-terminal was defined as variable modification. Carbamidomethylation on cysteines was set as a fix modification. The identified peptides were filtered using a FDR less than 5%. Peptide areas (obtained with Proteome Discoverer) were used for protein quantitation (Mascot decoy option activated).

Selection of target proteins, their peptides and transitions. Approximately 100 out of the 490 proteins had at least 1 hi-flyer peptide identified in the discovery scan. The Skyline analysis software was used to create spectral libraries and predict optimal collision energy, retention times, and transitions for these candidates.

Label-free selected reaction monitoring measurements

Optimization of Retention Times with Unscheduled Assays. Cleaned and unfractionated peptides were analyzed with a triple quadrupole mass spectrometer (AB SCIEX) run in unscheduled mode with retention time peptides to establish elution times. Peptides with high coefficients of variation across technical replicates were triaged.

Label-free quantitation of target proteins with scheduled assays. Having optimized elution times during unscheduled assay development, cleaned and unfractionated peptides for samples were run in scheduled mode on a triple quadrupole mass spectrometer (AB SCIEX). Approximately one-third of samples were run in biological triplicate, another one-third in biological duplicate, and the remainder one biological replicate was chosen randomly. Multiple technical replicates were also performed. Peptide quantitation was performed using Skyline software.

Data summarization (peptide-level to protein-level) and normalization (log2 drug:vehicle control. The R package MSstats v2.3.5 was used to explore the data, summarize multiple biological and technical replicates at the peptide-

level data to protein-level, and normalize drug treatment to vehicle treatment [Choi et al., 2014]. MSstats fits a mixed-effects linear model that partitions the variance in protein intensity into technical, biological, and experimental components [Clough et al., 2012]. To perform summarization, the dataProcess function was used with these arguments: logTrans = 2, normalization = constant, betweenRunInterferenceScore = FALSE. To perform normalization and statistical testing in comparison with DMSO vehicle treatment using the groupComparision function with arguments set as: labeled = FALSE, scopeOfBioReplication = restricted, scopeOfTechReplication = expanded, interference = FALSE, featureVar = TRUE, missing.action = nointeraction. In each drug condition, I set log2 ratios to 0 for proteins that had adjusted p-values greater than 0.05, i.e., insufficient evidence to reject the null hypothesis that its intensity is not different from drug vehicle.

Reverse-phase protein array measurements

note: these steps were performed by Xiaohong Jing in the Sander laboratory.
Cell pellets were thawed from -80°C to room temperature and lysed in CLB1 lysis buffer (Bayer Technology Services). Protein concentrations were determined by Pierce Coomassie Plus (Bradford) Assay Kit (LifeTechnologies). Protein concentrations were adjusted to 2 mg/ml in CLB1 lysis buffer and diluted 10x further in CSBL1 spotting buffer (Bayer Technology Services). A four-fold dilution series was prepared by the Biomek FXP Laboratory Automation Workstation (Beckman Coulter) to protein concentrations of 0.2, 0.15, 0.1, and 0.05 mg/mL. The dilution series was transferred to 384-well plates at 20 µL/well for spotting. Samples were spotted onto tantalum pentoxide-coated glass chips (ZeptoChip,

Bayer Technology Services) using the nano-plotter NP2.1 automatic pipetting systems (GeSiM).

Distance from pin to chip, setting piezo, and other spotting parameters were adjusted to achieve uniform, round, well-aligned spots. The piezoelectric tip of the nano-plotter aspirated several nanoliters and then deposited single 400 pL droplets onto the glass slides. A reference solution of BSA was also spotted adjacent to each dilution series. The spot layouts were not randomized and the experimenter was not blinded to the experimental conditions. There were six arrays per chip and each array was stained with one antibody or buffer control.

After spotting, chips were blocked for 20 minutes with an aerosol BSA solution using a customized blocking apparatus (ZeptoFog, Bayer Technology Services). Chips were washed in double-distilled H_2O and dried. Six spotted chips were inserted in a carrier (ZeptoCarrier, Bayer Technology Services). Antibodies were diluted in CAB1 following RPPA manufacturer recommendations. The chips were incubated with primary antibodies for 24 hours followed by incubation with a secondary antibody (Alex Fluor-647). Chips were washed with CAB1 and imaged with fluorescence optometry (ZeptoREADER, Bayer Technology Services) in the red channel (excitation 635 nM, emission 675 nM).

For each array four separate images were acquired using automatic exposure times from 1 to 10 seconds. Image analysis software (ZeptoView v.3.1) selected the best exposure. Normalization for each spot was performed against the BSA reference spots leading to a measurement of relative fluorescence intensity (RFI). Each spot was rated 'good, poor, or indeterminate' based on a linear change in signal intensity across the dilution series. 'Poor' and 'indeterminate' quality spots were removed from subsequent analyses. Antibodies

with a median intensity below 10x buffer control were considered noise and removed from subsequent analysis. Remaining data was normalized by double median normalization. In the first median pass, each antibody's RFI values were divided by the antibody's median RFI in all conditions. In the 2nd median pass, each condition's antibody RFI values were divided by the median RFI in the condition. The double-median normalized data were then summarized by taking the mean of three biological replicates.

2.3 Network analysis methods, predictions, and assessment

2.3.1 Nominating novel drug combinations with network analysis

To retrieve protein-protein interactions from a database involving all of the proteins measured by SRM and RPPA, antibody names had to first be mapped to gene symbols. In several cases, the mapping is not one-to-one and all gene symbols were retained. Phosphosites were given an edge to their corresponding protein. Interactions were retrieved from a protein-protein interaction database (Reactome, PathwayCommons2 v8) in simple interaction format. Proteins were removed from the protein-protein interaction network that did not show evidence of expression in the discovery phase data. To retrieve drug-protein interactions for FDA-approved drugs, I downloaded the supplementary files from Rask-Andersen et al [Rask-Andersen et al., 2011]. The statistical programming language R was used to merge and analyze the protein-protein:protein-drug network. For each perturbation I identified phosphosites and proteins with

strong responses. I then checked whether there is an FDA approved drug targeting the responder or its nearest neighbor yielding a hit list. Network druggability results were browsed visually using a network browser (Cytoscape v3.4.0). I performed manual literature curation of the hit list to identify studies involving the druggable nearest node and the drugs. I further filtered the results according to whether the drug is easily obtained commercially. I chose the top-9 ranked compounds for experimental testing. I estimated their IC50 doses in phenotypic assays from the literature shown in table 2.1.

Responder	Target	Drug	Dose	Reference
STMN1	aurora kinase B	barasertib	12.5 nM	[Zekri et al., 2015]
pRB1	CDK1/2/5	dinaciclib	50 nM	[Booher et al., 2014]
KLK3	PRKDC	enzastaurin	1.5 µM	[Graff et al., 2005]
YWHAZ	MAPK	losmapimod	10 nM	[Triantaphyllopoulos et al., 2010]
EIF4E	STAT3	napabucasin	1.5 µM	[Zhang et al., 2016]
KLK3	HDAC	panobinostat	5 nM	[Anne et al., 2013]
CDH1	SRC	saracatinib	500 nM	[Chang et al., 2008]
AIP	CSKN1A	silmitasertib	2.5 µM	[Ryu et al., 2012]
RHOA	ERK1/2	ulixertinib	100 nM	[Ward et al., 2015]

Table 2.1: Top-ranked predictions from druggable, nearest-responder-node analysis. References are given for the target or the drug. Doses are reported or estimated IC50 values from phenotypic assays.

2.3.2 Methods for testing network analysis predictions with phenotypic measurements of cell death

To test network analysis predictions, I assayed four of the original drug set (docetaxel, enzalutamide, MK2206, and temsirolimus) in all one-, two-, and three-drug combinations with the top-9 hits in the prediction drug set (barasertib, dinaciclib, enzastaurin, losmapimod, napabucasin, panobinostat, saracatinib,

silmitasertib, ulixertinib). To measure drug effects on cell death, LNCaP cells were seeded in 96-well plastic bottom plates at 12,500 cells per well in five biological replicates for all conditions. Cells were left at room temperature for 20 minutes for even seeding and placed in an incubator overnight. Drugs were administered concurrently at half the estimated IC50 value for all one-, two-, and three-drug combinations involving all one- and two-drug combinations of the original drug set: docetaxel 5 nM, enzalutamide 5 µM, MK2206 0.5 µM, temsirolimus 50 nM; with the prediction drug set: barasertib 12.5 nM, dinaciclib 50 nM, enzastaurin 1.5 µM, losmapimod 10 nM, napabucasin 1.5 µM, panobinostat 5 nM, saracatinib 500 nM, silmitasertib 2.5 µM, ulixertinib 100 nM. Drugs were used at the same concentration for all one-, two- and three-drug combinations. Drug dilutions were prepared such that the same volume of DMSO is added for each treatment and DMSO vehicle control. Five biological replicates were performed for each condition. The plate layouts were not randomized and the experimenter was not blinded to the experimental conditions to minimize confusion during manual labeling and pipetting.

To measure cell death, I used a reagent that fluoresces green when membrane integrity is lost (CellTox Green, Promega). Fluorescence readings were performed in a plate-reader (SpectraMax Gemini EM, Molecular Devices). Measurements were acquired at time 0 and 2 days after drug addition with these plate-reader settings: read mode: fluorescence top read; wavelengths: Ex 485, Em 530, auto-cutoff off; sensitivity: 30 readings per well, PMT auto; automix: 60 sec before read; auto-calibrate on; column wavelength priority: column priority. Instrument software (SoftMaxPro v5.4.1) was used to acquire and export data, which was further analyzed in R. Data were normalized by subtracting the time 0 background intensity for each well and summarized across five bi-

ological replicate wells by computing the mean and standard error. Further summarization was performed by comparison with drug vehicle DMSO.

2.3.3 Methods for testing network analysis predictions with molecular measurements by discovery proteomics

LNCaP prostate cancer cells (approximately 3 million) were harvested following drug treatments and resuspended in 800 µL of lysis buffer (8 M urea, 0.2% RapiGest, 0.1 M ammonium bicarbonate), vortexed for 10 s, and shaken at 1000 rpm at room temperature for 10 min. Samples were sonicated (amplitude 90%, cycle 0.6) for 2 x 2 min at 4°C and centrifuged at 13,200 rpm for 10 min. The clear lysate was transferred to a new tube, the amount of protein was determined using the BCA assay (Pierce, Thermo Scientific), and the protein concentration was diluted to 5 mg/mL with lysis buffer. The sample was reduced using TCEP (tris-2-carboxyethylphophine) and alkylated using iodoacetamide, and unreacted iodoacetamide was then neutralized with N-acetyl-cystein. The samples were then diluted with 0.1 M ammonium bicarbonate and 0.2% RapiGest to a final concentration of less than 2 M urea. Proteins were digested at a basic pH of 8.0 to 8.5 using sequence-grade modified trypsin (Promega) at a ratio of 1:100 for 14 h at 37°C. Digestion was stopped by decreasing the pH to less than 2 with formic acid. Samples were shaken another 30 min at 500 rpm at 37°C, followed by centrifugation at 13,200 rpm for 10 min to remove any insoluble fractions. Samples were cleaned with C18 columns.

Each sample was analyzed by LC MS/MS in an Orbitrip LTQ Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to nano-LC

(Proxeon, Odense, Denmark) equipped with a equipped with a reversed-phase chromatography 12-cm column with an inner diameter of 75 m, packed with 5 m C18 particles (Nikkyo Technos, Japan). Chromatographic gradients started at 97% buffer A and 3% buffer B with a flow rate of 300 nL/min and gradually increased to 93% buffer A and 7% buffer B in 1 min and to 65% bufferA and 35% buffer B in 60 minutes. After each analysis, the column was washed for 10 min with 10% buffer A and 90% buffer (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in acetonitrile). The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.2 kV and source temperature at 250 C. Ultramark 1621 for the FT mass analyzer was used for external calibration prior the analyses. Moreover, an internal calibration was also performed using the background poly- siloxane ion signal at m/z 445.1200.

The instrument was operated in DDA mode and full MS scans with 1 microscan at resolution of 60,000 were used over a mass range of m/z 2502000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1e6, dynamic exclusion (60 s with a repeat duration of 30 s [repeat count = 1], and an exclusion list size of 500 ions), and the charge state filter disqualifying singly charged peptides was activated. Following each survey scan, the top 20 most intense ions with multiple charged ions above a threshold ion count of 5000 were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision- induced dissociation (CID) were acquired in the linear ion trap, AGC was set to 5e4, isolation window of 2.0 m/z, and activation time of 0.1 ms, and maximum injection time of 100 ms was used. All data were acquired with Xcalibur software v2.2.

Acquired data were analyzed using the Proteome Discoverer software suite

(v1.3.0.339, Thermo Fisher Scientific), and the Mascot search engine (v2.3, Matrix Science) was used for peptide identification. Data were searched against an in-house generated database containing all proteins corresponding to human in the Swissprot human database (as of 3/2013) with 600 added common contaminants, with a total number of 37,694 sequences in the database. A precursor ion mass tolerance of 7 ppm at the MS1 level was used, and up to three miscleavages for trypsin were allowed. The fragment ion mass tolerance was set to 0.5 Da. Oxidation of methionine and protein acetylation at the N-terminal was defined as variable modification. Carbamidomethylation on cysteines was set as a fix modification. The identified peptides were filtered using a FDR less than 5%. Peptide areas (obtained with Proteome Discoverer) were used for protein quantitation (Mascot decoy option activated). Analysis of enriched pathways was performed using the GSA geneset analysis method, which uses the maxmean statistic to compute an enrichment score for collections of gene sets [Efron and Tibshirani, 2007]. Here the Reactome genesets were used.

CHAPTER 3

RESULTS - QUANTIFYING MOLECULAR AND PHENOTYPIC RESPONSES TO DRUG PERTURBATIONS

3.1 Establishment of SRM assays

An initial protein list of interest in prostate cancer (PLIPCa) consisting of 490 candidates was developed through manual literature curation. These proteins proved to be challenging targets. The top left panel of figure 3.1 compares the abundances of all proteins detected in LNCaP by discovery MS following extensive offline fractionation in Geiger et al. [Geiger et al., 2012] against those detected in our discovery MS from strong cation exchange fractionation and those within the candidate list of 490 proteins. The result of this comparison is that the majority of the 490 list are undetectable in all datasets; moreover, those that are detected are highly enriched for low abundant proteins. The left bottom panel shows that most proteins in the final target list were detected in either the strong cation exchange or isoelectric focusing fractions, with a few peptides originating from unfractionated shotgun discovery scans or from a peptide database of hi-fliers [Kusebauch et al., 2016]. The middle panel of figure 3.1 shows elution times and transitions for several examples in the target list. The right panel shows that our selection of the 24 hour time point is reasonable.

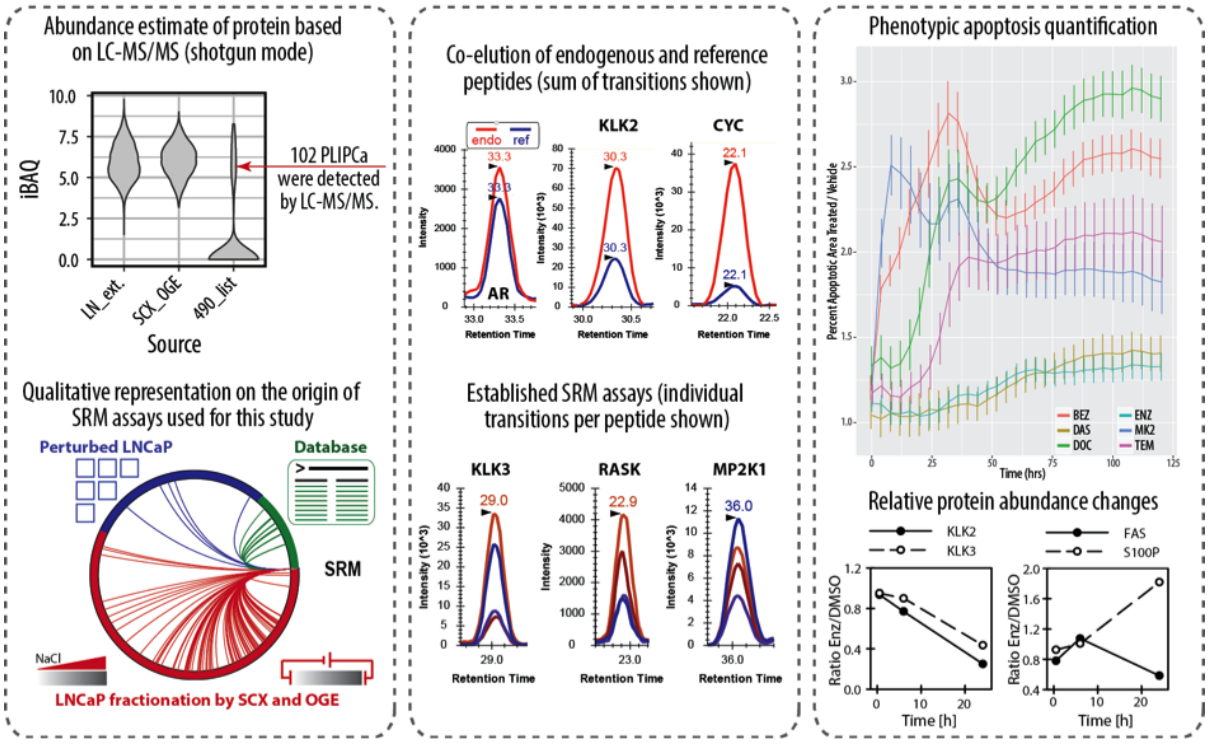

Figure 3.1: **SRM assay development and exploratory perturbation for time point selection.**
Left panel. Violin plot showing protein abundance distribution of proteins quantified in LNCaP cells from Geiger et al., this study and estimated abundance values of the target PLIPCa list. Circular Interaction Graph for Proteomics representing the origin of all peptides corresponding to the target PLIPCa proteins.
Middle panel. SRM assay development. Top row showing sum of transitions per peptide for three proteins (AR, KLK3 and CYC). Endogenous peptides co-elute with stable isotope labelled reference peptides. Bottom row showing individual transitions per peptide which are base line separated.
Right panel. Time point selection. Top: six drugs were chosen and automated fluorescent microscopy quantified for apoptosis as a function of time. Different drugs appear to trigger apoptosis with varying kinetics. Therefore, our choice of the 24 hour timepoint is arbitrary and misses the peak signal for many conditions. Bottom: the proteome of LNCaP cells treated with enzalutamide was quantified at 0.5, 6 and 24 h post treatment. The relative abundance changes for four proteins are shown.

3.2 Phenotypic and proteomic responses to perturbations

Targeted proteomics measurements included 52 proteins measured by SRM and 70 (phospho)-proteins measured by RPPA quantified across 6 one-drug and the 15 possible two-drug combinations, plus a drug vehicle control. Data summarized across biological and technical replicates and summarized as log2 ratio to drug vehicle is show in figure 3.2. Most protein levels do not increase or decrease very much following drug perturbation. Some (phospho)-proteins are consistently increased or decreased relative to drug vehicle, such as YWHAE, YWHAB, and PAK1. Proteins showing relatively high variation among drug conditions include pRB, EIF4E, and ACTR2. Measurements of pAKT in AKTi conditions confirm known drug effects. The SRC inhibitor dasatinib has noticeably weaker effects on (phospho)-protein levels than other drugs at the doses assayed.

3.3 Network analysis and predictions

A simple, binary network analysis method shown in figure 3.3 was used to identify candidate markers of response and resistance. For each drug condition, if a (phospho)-protein increased or decreased twice that of vehicle treatment, it was retained as a strong responder. These strong responders were mapped onto a protein-protein-drug interaction network, where the drugs have been filtered for commercial availability. Strong responders with druggable nearest neighbors are nominated to enhance or counteract mechanisms of response or resistance, respectively. In all cases, the strong responder was not located immediately upstream or downstream of the drug target, but was located instead

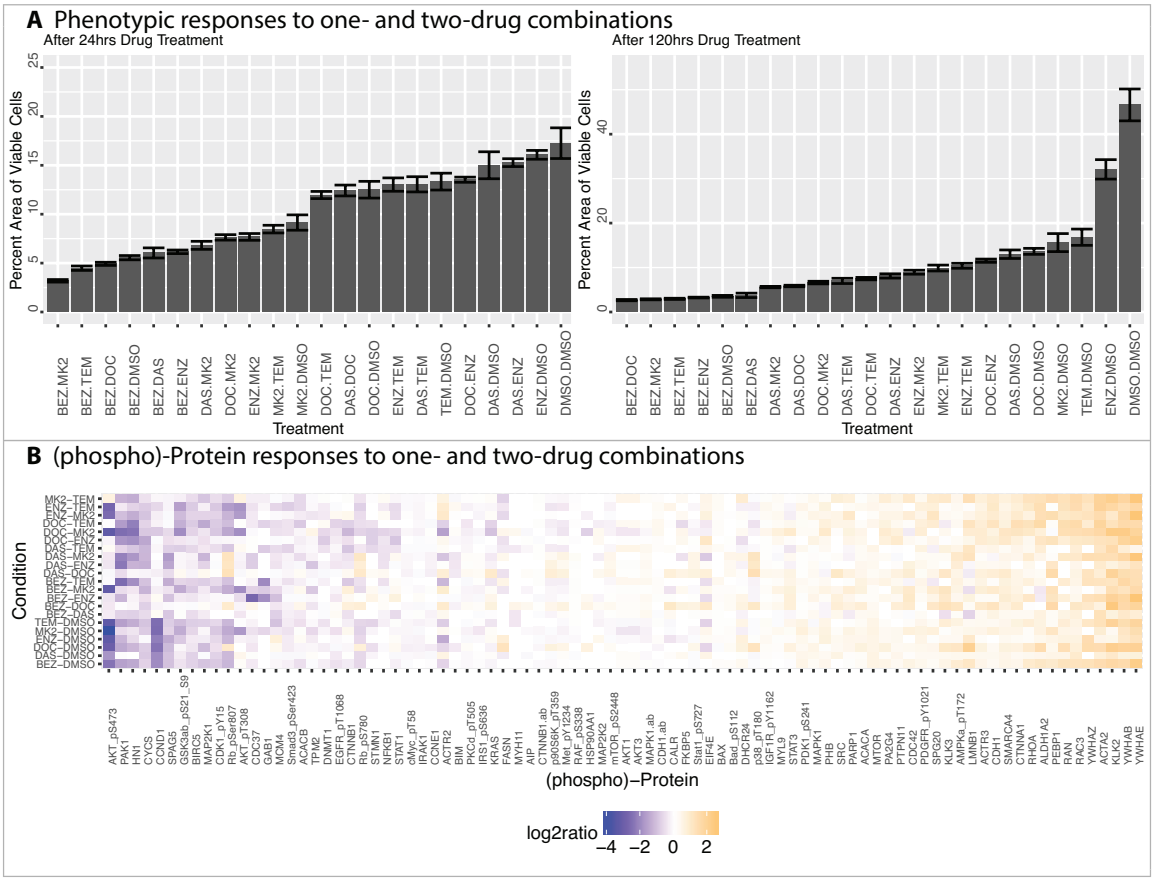

Figure 3.2: **Phenotypic and proteomic responses to one- and two-drug treatments.**
(A) Phenotypic effects of one- and two-drug combinations administered concurrently at single doses were assessed by microscopy. Drug names, targets and doses are shown in Table 1. Drug effects were quantified by percent area of viable cells, which is computed by subtracting the area of fluorescence channel signal of the apoptosis probe (caspase 3/7 activation) from the phase contrast channel measurement of cell area. Cells grew in clumps that could not be segmented into individual cells for a computation of percent apoptotic cells. Comparing drug effects at day 5 (right panel) to day 1 (left panel) reveals that effects are more pronounced and more varied at day 5 than day 1. Error bars are +/- SEM.
(B) Integrated log2 responses of targets compared to vehicle 24 hours after drug treatment. Proteins were measured by selected-reaction monitoring mass spectrometry (SRM) and additional (phospho)-proteins were measured by reverse-phase protein arrays (RPPA).

in a parallel pathway. It is possible that greater coverage of signaling pathways through additional SRM or RPPA targets might lead serial pathway inhibition strategies.

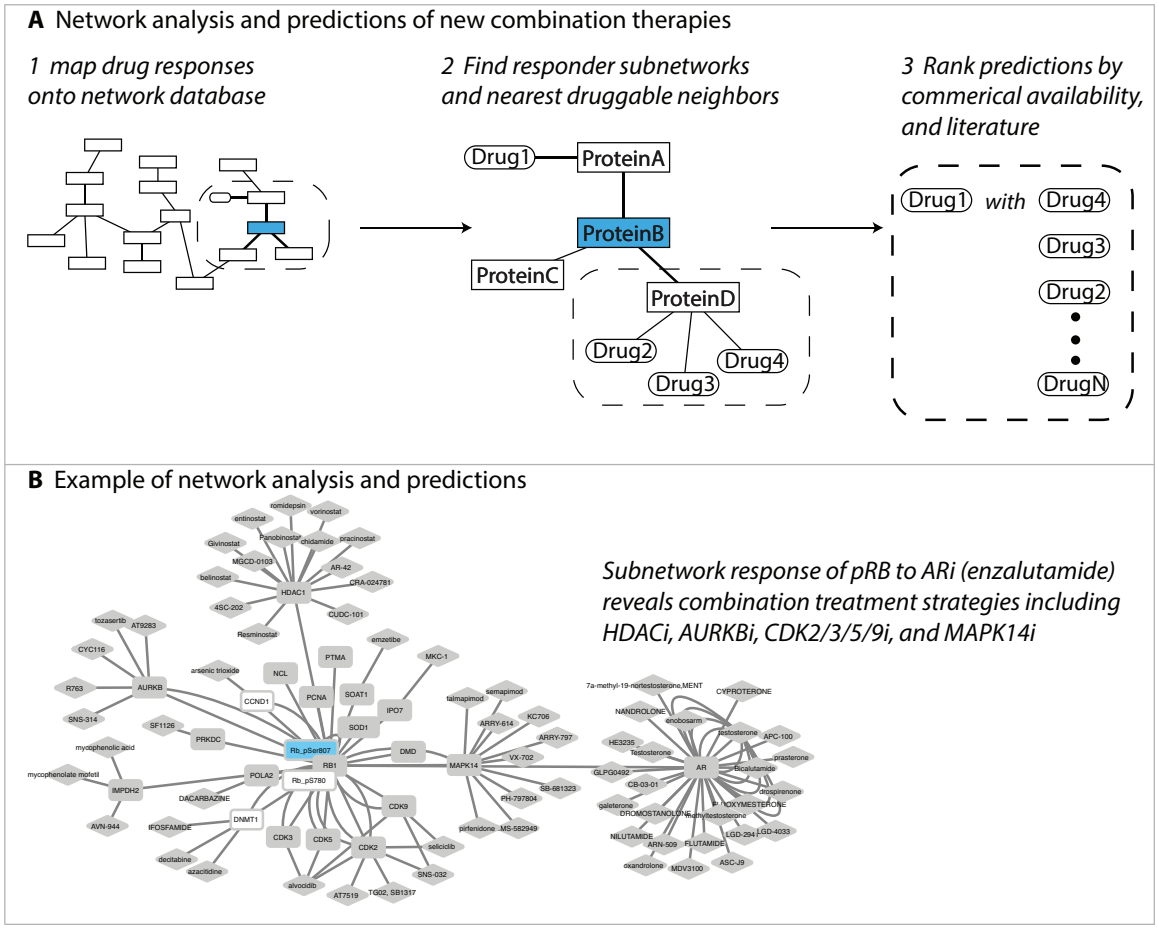

Figure 3.3: **Network analysis of drug treatment responses identifies potential network vulnerabilities and new drug combination strategies.**
(A) The analysis method has three steps: (1) drug response data are mapped onto a database network of protein-protein and protein-drug interactions from PathwayCommons.org; (2) nearest-neighbor subnetworks around proteins with strong responses to perturbations are extracted for visualization and further analysis; (3) an additional filter is applied for commercial availability and the extent of literature on the role of the target and/or efficacy of the drug in prostate cancer because there are typically numerous options for drugging the nearest node to a (phospho)-protein with strong responses.
(B) Phospho-RB has strong responses to numerous perturbations including AR inhibition with enzalutamide. This subnetwork shows that there are no available drugs for targeting pRB directly but several options for drugging its nearest neighbors.

CHAPTER 4

RESULTS & DISCUSSION - STRATEGY 1: CO-TARGETING PI3K SIGNALING AND CELL CYCLE

4.1 Testing predictions uncovers a PI3K signaling & cell cycle inhibition strategy

To test network analysis predictions I assayed four of the original drug set (docetaxel, enzalutamide, MK2206, and temsirolimus) in all one-, two-, and three-drug combinations with the top-10 predictions from network analysis: (barasertib, BV02, dinaciclib, enzastaurin, losmapimod, napabucasin, panobinostat, saracatinib, silmitasertib, ulixertinib). Drugs were administered concurrently at half the estimated IC50 value for all one-, two-, and three-drug combinations involving all one- and two-drug combinations of the original drug set. To measure cell death I used a reagent that fluoresces green when membrane integrity is lost. Only CDK1/2 inhibitor dinaciclib yielded effective two- and three-drug combinations, shown in figure 4.1.

4.2 Discussion on PI3K signaling & cell cycle inhibition strategy

Cell cycle inhibitors are considered a promising strategy for many cancers [Lapenna and Giordano, 2009]. Stice and colleagues propose CDK4/6 inhibitors as alternatives to taxane therapy for prostate cancer [Stice et al., 2017].

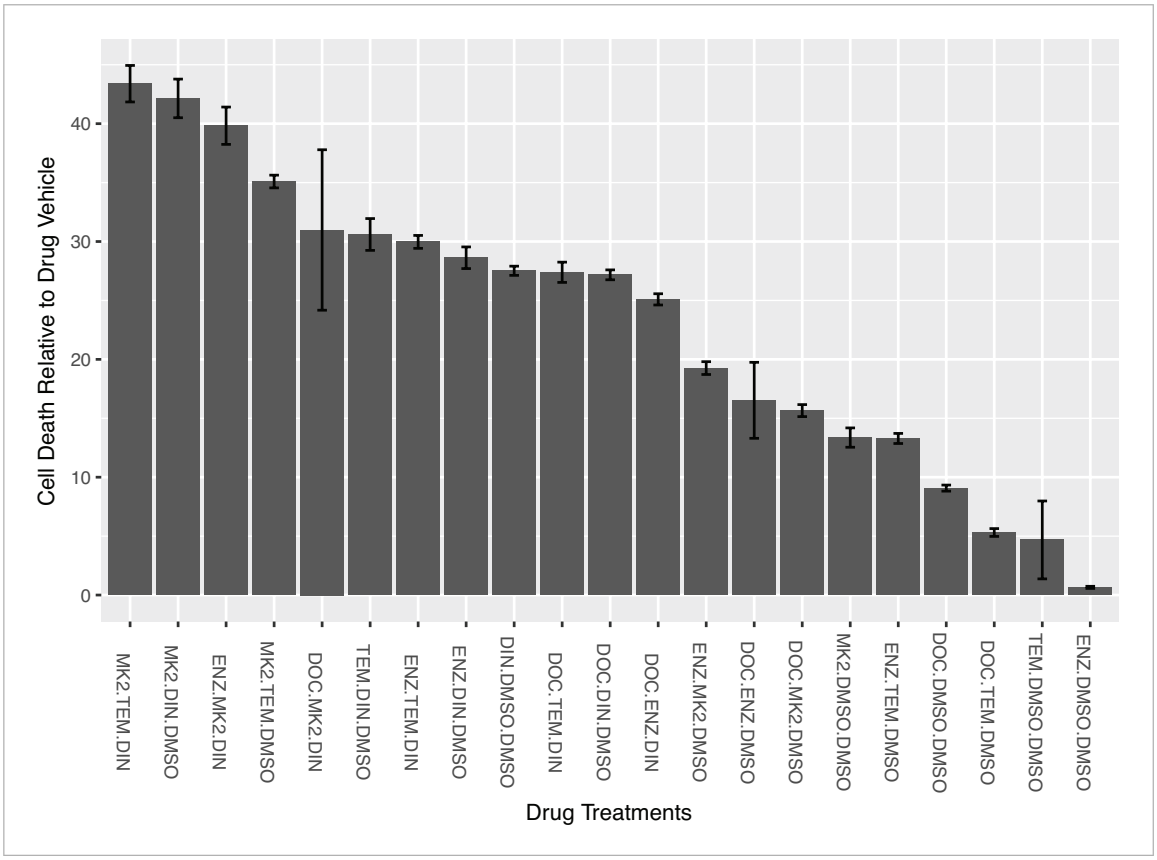

Figure 4.1: **Testing network analysis predictions of the efficacy of CDK inhibition in a cell death assay.** The CDK1/2/5 inhibitor dinaciclib was tested in all one-, two-, and three-drug combinations co-administered at single doses with 4 members of the first drug set: docetaxel (microtuble inhibitor), enzalutamide (AR inhibitor), MK2206 (AKT inhibitor), and temsirolimus. Cell death was assessed after 48 hours of drug treatments. Doses are given in Table 2. Dinaciclib is most effective in combination with MK2206 and MK2206 with temsirolimus. Dinaciclib is relatively effective on its own. Error bars are +/- SEM. Nine additional drug predictions from network analysis were assayed similarly in all one-, two-, and three-drug combinations but seven did not show efficacy at the dose tested and for two drugs effects could not be determined.

Recent studies in prostate cancer biology have found that CDK5 acts as a critical signaling hub driving AR and AKT signaling, releasing cell cycle breaks [Lindqvist et al., 2015]. A similar study found CDK5 regulates both STAT3 and AR [Hsu et al., 2013]. In a review article, Balk and Knudsen describe feedback

loops operating between AR, mTOR and cell cycle proteins, which are illustrated in figure 4.2 [Balk, 2007]. Gao et al. also reviewed the literature and concluded similar connectivity between mTOR and cell cycle progression at the G1/S transition [Gao et al., 2016].

Dinaciclib is a highly selective and potent inhibitor of CDK1/2/5 [Paruch et al., 2010]. It has been found to synergize with AKT inhibitor MK2206 in pancreatic cancer [Hu et al., 2015] and AKT inhibitors in ovarian cancer [Au-Yeung et al., 2017]. In figure 4.2, I propose a highly simplified model where dinaciclib enhances the cell cycle arrest response caused by AR inhibition with enzalutamide and AKT inhibition with MK2206. This can be considered a serial pathway inhibition strategy because blocking mTOR via AKT lowers CDK4/6-mediated G1/S progression and dinaciclib blocks S/G2 and G2/M progression by inhibiting CDK2 and CDK1, respectively. Additional experiments measuring cell cycle status and (phospho)-protein status of RB, CDK4/5, CDK1, and CDK2 are necessary to test the hypothetical mechanism in figure 4.2.

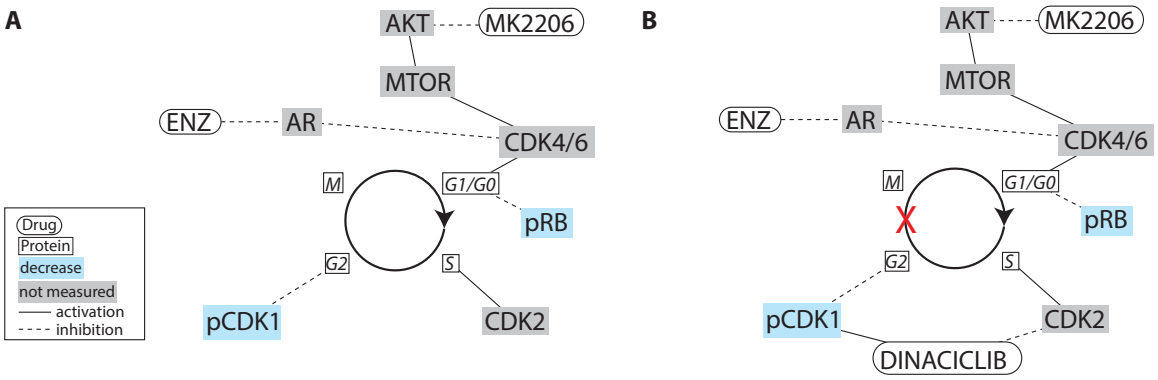

Figure 4.2: **Combined inhibition of AKT and cell cycle is more effective than either drug alone.** Following single drug treatment with PI3K pathway inhibitors network analysis of targeted proteomics measurements of drug responses led to the hypothesis that augmenting the response of phospho-RB by inhibition of CDK1/2 would enhance drug efficacy. Single drug efficacy may be limited by a feedback loop operating between MTOR and PI3K or between AR and CDK4. A drug targeting CDK1/2 (dinaciclib DIN) was chosen based upon commercial availability, status in clinical trials, and potency. A cell death assay confirmed the prediction. This inhibition strategy is similar to serial pathway inhibition, however, the effects are probably more complicated because of the many interactions among these and other cell cycle proteins not shown.

CHAPTER 5

RESULTS & DISCUSSION - STRATEGY 2: CO-TARGETING 14-3-3 PROTEINS IN COMBINATION WITH AKT AND AR

5.1 Testing predictions uncovers a 14-3-3 PI3K, and AR inhibition strategy

During network analysis I observed that 14-3-3 paralogs YWHAB, YWHAE, and YWHAZ increased following the majority of drug perturbations. Qualitative network analysis predicted that 14-3-3 protein increases might be general mechanisms of resistance to drug perturbation. My interest in testing this prediction was augmented when I noticed that 14-3-3-zeta (YWHAZ) was recurrently amplified in 27% of metastatic tumors shown in figure 5.2. To test network analysis predictions, I assayed two of the initial drugs that have the most recurrent alterations in metastatic tumors: AR inhibitor enzalutamide and AKT inhibitor MK2206 with 14-3-3 inhibitor BV02. Drugs were administered concurrently at the estimated IC50 value BV02, 10 μM enzalutamide, and 500 nM MK2006 in combinations: BV02 alone, BV02 + MK2206, BV02 + enzalutamide, and BV02 + MK2206 + enzalutamide. Cell proliferation was assessed by a colorimetric endpoint assay using tetrazolium dye MTT, which is converted to formazan by oxidoreductase enzymes in metabolically active cells [van Meerloo et al., 2011]. I also performed discovery proteomics on each perturbation condition using an orbitrap mass spectrometer to measure the molecular effects of drug perturbation.

Analysis of drug response phenotypes revealed the triple combination of

ARi+AKTi+YWHAZi to effect numerous proteins, including many from the target list shown in figure 5.3, with DNA-repair protein BRCA1 and transcription factor NKX3.1 showing large decreases compared to drug vehicle. Geneset analysis identified decreases in multiple pathways relating to cell cycle and DNA repair shown in figure 5.4. In the phenotypic assay of cell proliferation, the triple combination showed increased efficacy over the one- and two-drug combinations tested, shown in figure 5.5. These experiments are preliminary because not all one- and two-drug combinations of the drugs were assayed. In particular, without measurements made for the AKTi+ARi combination, efficacy of the triple combination of AKTi+ARi+YWHAZi may be less significant. Experiments to measure all one- and two-drug combinations are being performed by H. Alexander Ebhardt at his laboratory in University College, Dublin.

5.2 Discussion on 14-3-3, PI3K, and AR inhibition strategy

Molecular and phenotypic evidence indicates that triple inibition of 14-3-3, PI3K and AR is a promising strategy. Drug response data in figure 5.4 indicate that the triple combination appears to lower DNA repair genes suggesting that AR activity is lowered because AR has been shown to coordinate DNA repair and radioresistance [Polkinghorn et al., 2013]. Moreover, figure 5.5A shows the the triple combination lowers expression of FKBP5, an AR target gene that mediates reciprocal feedback inhibition between PI3K and AR pathways [Carver et al., 2011]. Interestingly this pattern is reversed in PC3; however, combined levels of 14-3-3 are low in both cell lines (figure 5.5A). These results provide a number of reasons for continued evaluation of the role of 14-3-3 proteins in prostate cancer and their potential as drug targets.

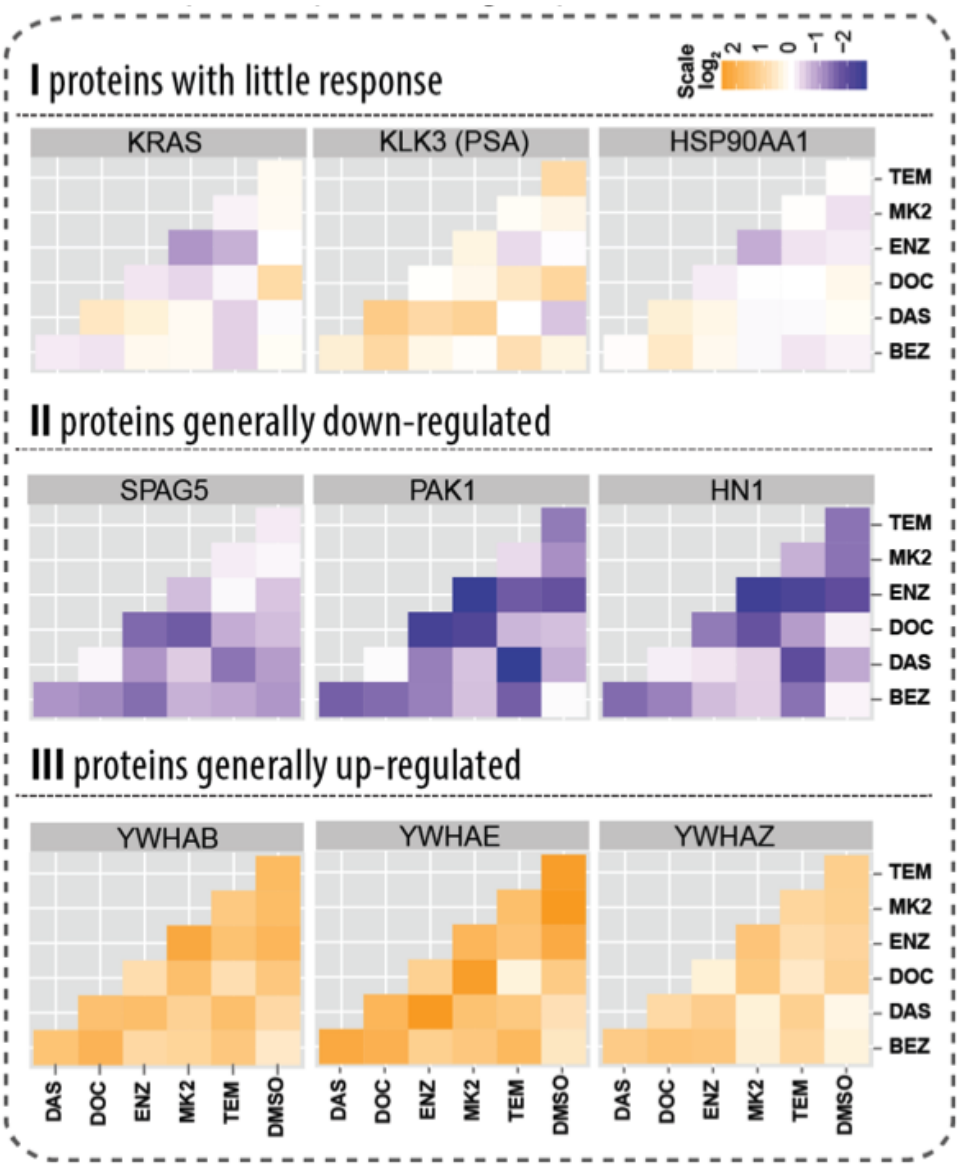

Figure 5.1: **Patterns of protein responses across perturbations identify consistent strong responders.**
I. Proteins frequently quantified in prostate cancer (e.g. KLK3 [PSA] or HSP90) show relatively small changes across pharmacologically perturbed conditions.
II. Proteins consistently decreasing as a function of perturbation. PAK1 and HN1 have not been shown to be important for prostate cancer.
III. Proteins with consistent increases following drug perturbation include 14-3-3 proteins.

The seven paralogous 14-3-3 proteins in humans have been considered as potential therapeutic targets [Zhao et al., 2011]. 14-3-3 proteins are highly conserved and recognize phospho-serine binding motifs RSXpSXP (mode 1) and RXY/FXpSXP (mode 2) [Fu et al., 2000, Aitken, 2006]. They interact with more than 100 proteins and play regulatory roles including alteration of enzymatic activity, alteration of DNA-binding activity, sequestration, altering protein-protein interactions, and adaptor protein functions [Hermeking, 2003]. In particular their roles in inhibiting apoptosis makes them of interest as therapeutic targets [Hermeking, 2003, Aghazadeh and Papadopoulos, 2016].

In prostate cancer, 14-3-3 protein YWHAZ was shown by Menon and colleagues to be amplified in 48% of CRPC patients and knockdown reduced proliferation and migration *in vitro* [Menon et al., 2013]. In addition to 14-3-3 proteins' roles in inhibiting apoptosis, Oh and colleagues found that non-sigma 14-3-3 proteins protect ETV1 from degradation, thereby promoting prostate tumorigenesis [Oh et al., 2013]. Ruenauver and colleagues examined the prognostic significance of YWHAZ and found high expression is associated with high gleason score, higher risk of CRPC development, and reduced survival time [Rüenauver et al., 2014].

During our investigation of YWHAZ in prostate cancer I realized that it is located on 8q, which is a recurrent whole-arm chromosomal amplification in prostate cancer and has prognostic significance [Silva et al., 2016]. cMYC is considered a driver on 8q and additional genes have been shown to play roles including AR-coactivator NCOA2 and YWHAZ [Silva et al., 2016, Menon et al., 2013]. Intriguingly, I also observe recurrent co-occurrence of nMYC and YWHAQ amplification on 2p (figure 5.5B). Our result that triple

combination AKTi+ARi+YWHAZi causes decreases in cell cycle and DNA repair proteins suggests an inhibition of AR activity because AR is known to master regulate DNA repair [Polkinghorn et al., 2013]. This evidence from the literature taken together with our preliminary findings that YWHAZ may play a role in resistance mechanisms to drug treatment are plausible and warrant further investigation. Inhibitor BV02 was developed as a c-Abl and 14-3-3σ inhibitor and its specificity for the other paralogs is unclear [Mancini et al., 2011], so siRNA experiments are currently being performed by H. Alexander Ebhardt at his laboratory at University College, Dublin.

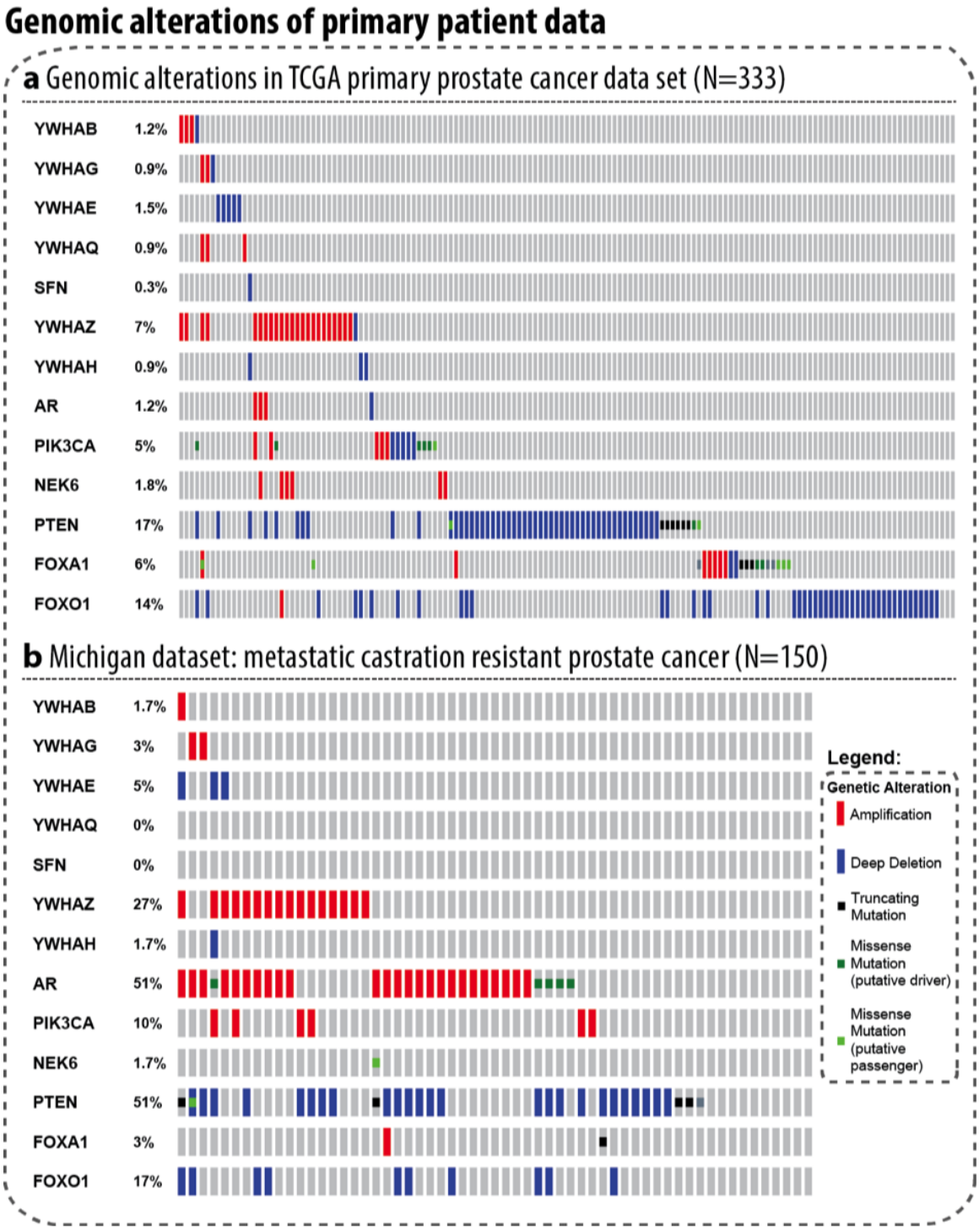

Figure 5.2: **Recurrent alterations in primary and metastatic tumors show amplifications in 14-3-3 protein YWHAZ.**
(A) Oncoprint of primary prostate tumor data shows recurrent alterations in PTEN and FOXO1.
(B) Oncoprint of metastatic prostate tumor patient data shows recurrent amplifications in AR and YWHAZ. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]]

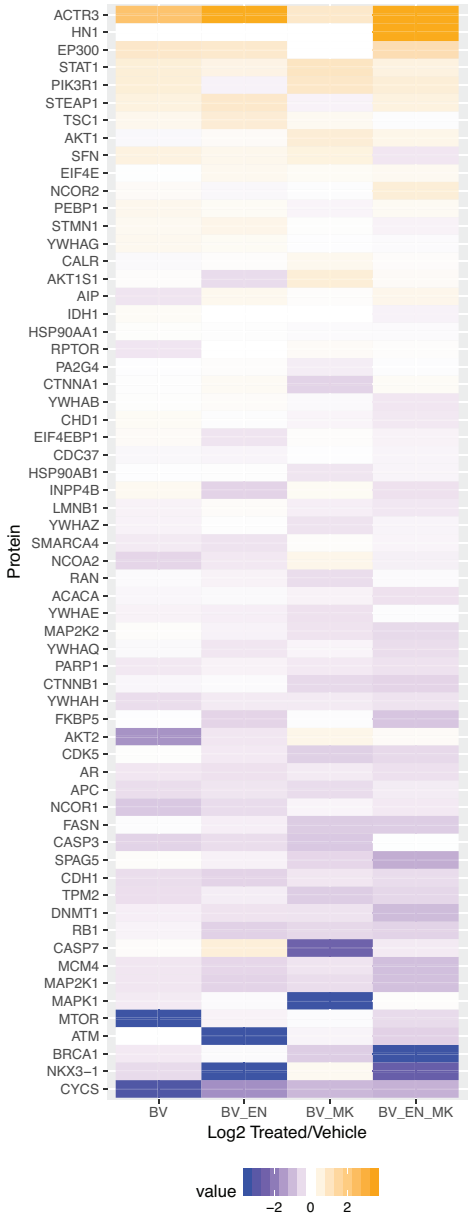

Figure 5.3: **Effects on relative protein expression of 14-3-3 inhibitor BV02 for proteins in the target list.**

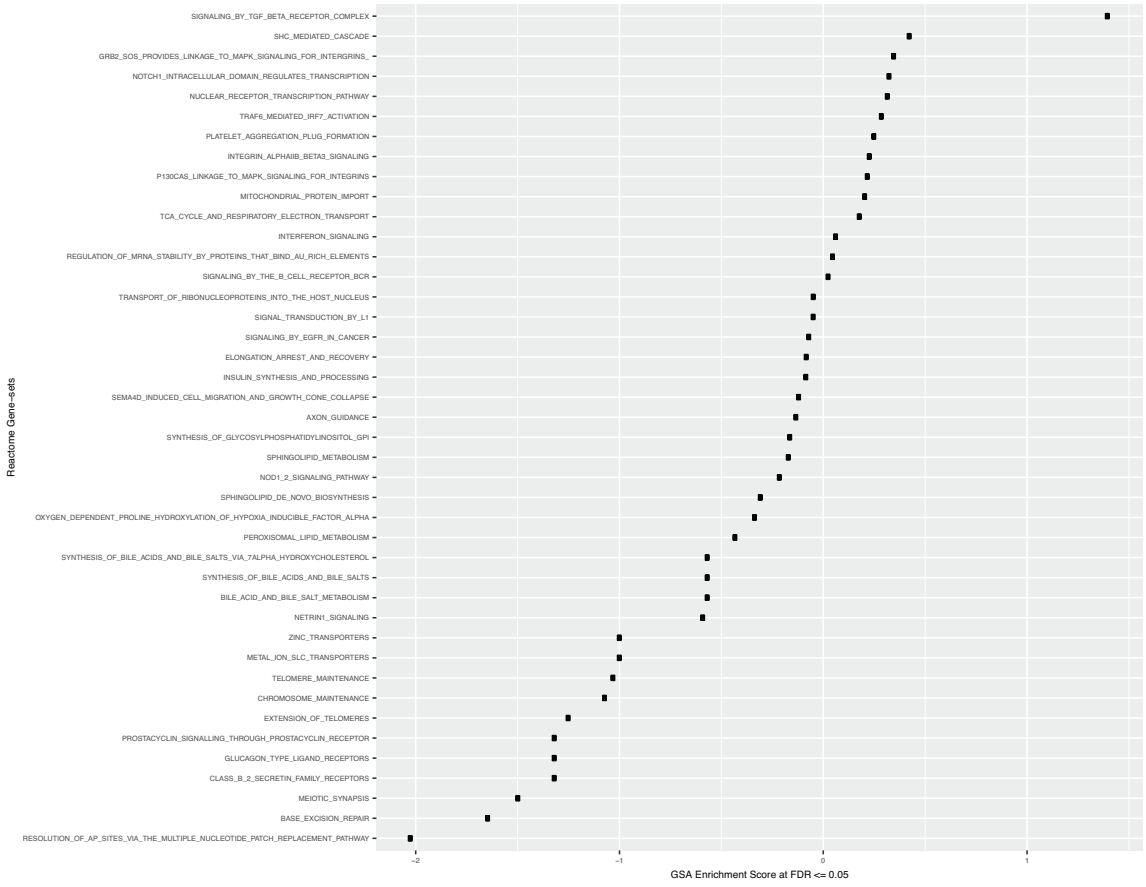

Figure 5.4: **Geneset analysis of triple combination therapy of ARi, AKTi, and YWHAZi indicates decreases in DNA repair and other processes.**

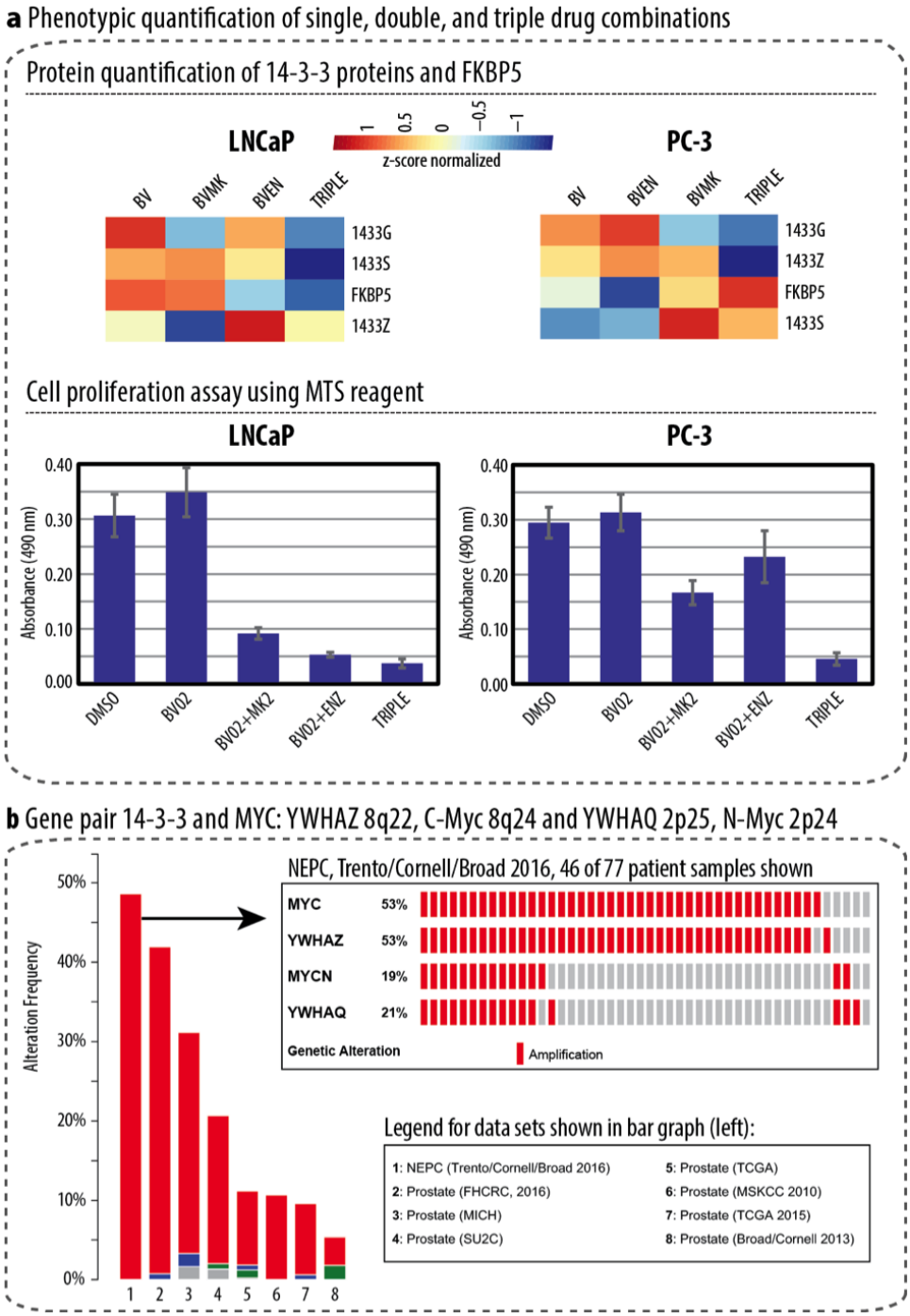

Figure 5.5: **Combination therapy targeting YWHAZ, AKT and AR and patient alterations in YWHAZ.**
(A) Molecular and phenotypic effects of YWHAZ inhibitor BV02
(B) YWHAZ amplifications co-occur with MYC on 8q or 2p. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]]

CHAPTER 6

ALTERNATIVE APPROACH: IDENTIFYING ANDROGEN RECEPTOR INTERACTING PROTEINS

6.1 Introduction - AR-interacting proteins may be good drug targets

AR signaling is necessary for prostate tumorigenesis and its blockade by androgen deprivation therapy results in the emergence of castrate resistance, that is treated by second line blockade with AR and CYP17A1 inhibitors to which resistance again emerges [Spratt et al., 2016]. Given this importance of AR I sought to understand how co-activators, co-repressors, and other interactors influence its signaling in prostate cancer.

AR-interacting proteins have been previously identified in several studies. Jasavala and colleagues used the 293 human embryonic kidney cell line to identify interactors in cytosolic and nuclear fractions using AP-MS [Jasavala et al., 2007]. Norris and colleagues used a high-throughput T7 phage display to identify ligand-bound interactors [Norris et al., 2009]. The effect of AR mutations WT, T877A, and 0CAG on interactors was examined in monkey kidney fibroblast-like COS-1 cells following mibolerone using his-tagged AP-MS [Paliouras et al., 2011]. A variation of this approach developed an N-TAP-mAR in engineered P17 and M7 mouse epididymal cells and showed loss of only a few interactors in the tagged version in proliferating (37°C) and nonproliferating (33°C) conditions in nuclear and cytosolic fractions by AP-MS [Mooslehner et al., 2012]. Effects of the T877A mutations were also exam-

ined in LNCaP cells under eight treatment conditions (dihydrotestosterone, mibolerone, R1881, testosterone, estradiol, progesterone, dexamethasone, and cyproterone acetate) by AP-MS and differences in interactors were shown to have prognostic value for survival in patient cohorts [Zaman et al., 2014]. AR interactors are also catalogued in databases, e.g., PathwayCommons.org [Cerami et al., 2011]. Despite these resources, there is a gap in our understanding of how clinically-relevant perturbations such as enzalutamide treatment affect AR interactors in prostate cancer.

I aimed to identify AR interactors in the VCaP cell line following perturbation with AR ligand DHT, agonist bicalutamide, antagonist enzalutamide, and drug vehicle DMSO [Tran et al., 2009]. I chose VCaP because it has relatively high expression of AR protein [van Bokhoven et al., 2003]. Our experimental design included subcellular fractionation to identify interactors in cytosolic and nuclear-enriched fractions because AR shuttles between cytoplasm and nucleus upon activation and its translocation is affected by enzalutamide [Tran et al., 2009]. One approach to identifying protein interactors is affinity-purification mass spectrometry (AP-MS) [Gstaiger and Aebersold, 2009]. By quantifying ”interactomes”, AP-MS has made important contributions to network biology, including metabolism [Rouleau et al., 2017] and nuclear receptor studies including the estrogen receptor [Mohammed et al., 2013].

Our AP-MS workflow is shown in figure 6.1. There are a variety of methods for performing affinity-purification, e.g., FLAG-tagging and I chose to use a polyclonal antibody against the N-terminal region, shown in figure 6.1. All affinity-purification methods have the potential to disrupt protein-protein interactions. By choosing an antibody that recognizes an epitope within the first

50 amino acids at the N-terminus this issue may be minimized, see figure 6.1B. After identifying AR interactors I will analyze how they compare to known AR interactors in databases, how they change with ligand, agonist, and antagonist perturbations, and how they are recurrently altered in patient cohorts, see figure 6.1A.

6.2 Methods - identification of AR-interacting proteins in a cellular context relevant to prostate cancer with AP-MS

note: Wetlab experiments were performed by Wassim Abida in the Sawyers laboratory and proteomics by H. Alexander Ebhardt in the Aebersold laboratory. I analyzed the data and created the figures. Appendix B describes authorship in more detail.

6.2.1 Cell culture and drug treatments

The VCaP cell line is AR sensitive and was isolated from a vertebral bone metastasis from a patient with CRPC [van Bokhoven et al., 2003]. Cultures were maintained in a humid 5% CO_2 atmosphere at 37°C and grown in DMEM medium supplemented with 10% heat-inactivated fetal bovine serum, 100 units/ml each of penicillin and streptomycin, and 2mM L-glutamine. Cells were treated with 1nM DHT, 10µM bicalutamide, 10µM enzalutamide, and DMSO for 2 hours.

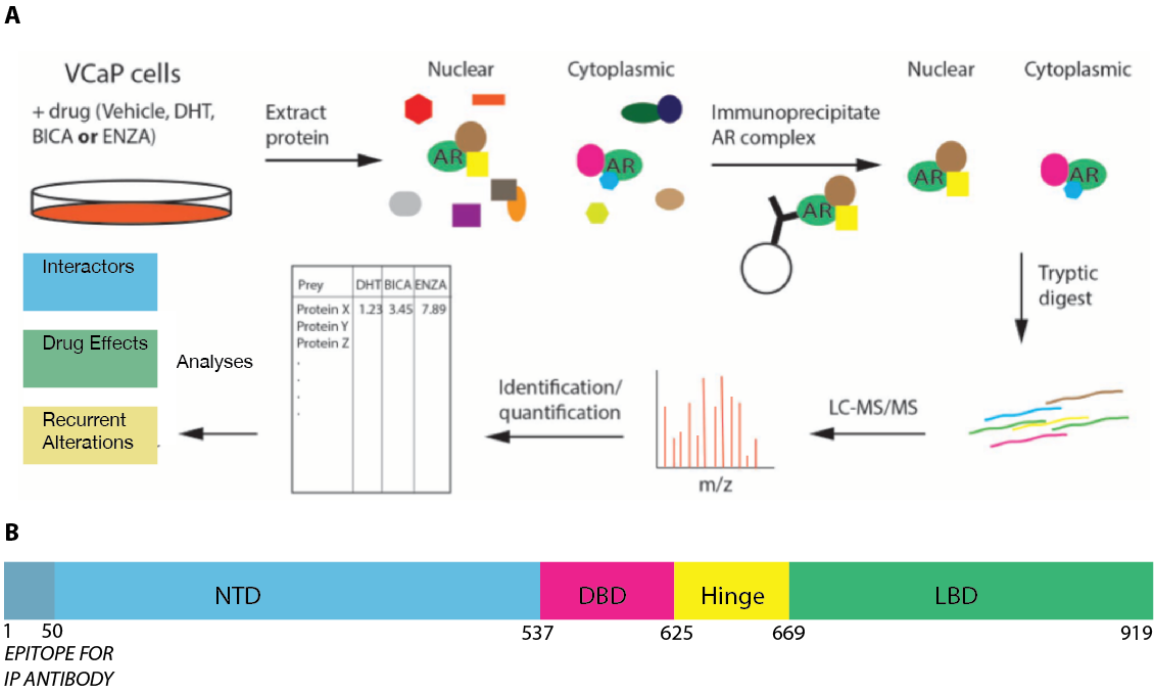

Figure 6.1: **Identification of AR-interacting proteins after ligand, agonist, and antagonist treatment by affinity purification mass spectrometry (AP-MS).**
(A) The prostate cancer cell line VCaP is an early metastatic model that expresses full-length AR and is sensitive to its inhibition. VCaP cells were grown in normal culture media and treated with ligand dihydroxytestosterone (DHT), agonist bicalutamide (BICA), and antagonist enzalutamide (ENZA). Cells were lysed and fractionated into cytoplasmic- and nuclear-enriched. A polyclonal antibody recognizing the N-terminal domain of AR immunoprecipitated interacting proteins. Analysis by mass spectrometry revealed a number of proteins differentially abundant across treatments. Additional analyses integrating databases of AR-interacting proteins, recurrent alterations in prostate cancer patients, and druggability prioritized interactors for additional experimentation.
(B) Domain structure of full-length androgen receptor including the N-terminal domain (NTD), DNA-binding domain (DBD), hinge, and ligand-binding domain (LBD). The LBD is thought to be the site of drug binding. The antibody used for pull-down is Santa Cruz N20 binds in the first 50 amino acids potentially disrupting some protein-protein interactions. Other methods, such as FLAG- or HIS-tags would also suffer from this problem.

6.2.2 Affinity Purification Mass Spectrometry

AP-MS methods follow Hauri and colleagues with some modifications [Hauri et al., 2013]. Cells were collected, frozen in liquid nitrogen and stored at 80°C prior to protein complex purification. The frozen cell pellets were re-suspended in 4 ml lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 50 mM NaF, 0.5% Igepal CA-630 (Nonidet P-40 Substitute), 200 mM sodium orthovanodate, 1 mM phenylmethane sulfonylfluoride, 20 mg/ml Avidin and 1x Protease Inhibitor mix (Sigma) and incubated on ice for 10 min. Insoluble material was removed by centrifugation. Complexes were pulled down by immunoprecipitation with the N20 anti-AR antibody (SantaCruz). IgG only served as a negative control. Cysteine bonds were reduced in the dark with 5 mM TCEP for 30 min at 37°C and alkylated in 10 mM iodoacetamide for 30 minutes at room temperature. Samples were diluted with ammonium bicarbonate to 1.5 M urea and digested with 1 µg trypsin (Promega) overnight at 37°C. The peptides were purified using C18 microspin columns (Nest Group) in 0.1% formic acid, 1% acetonitrile.

Discovery mass spectrometry measurements were performed on an Orbitrap XL (Thermo Fisher Scientific). Peptide separation was performed by a Proxeon EASY-nLC II liquid chromatography system (Thermo Fisher Scientific) connected to an RP-HPLC column packed with Magic C18 AQ resin (WICOM International). Solvent A was used as RP-HPLC stationary phase (0.1% formic acid, 2% acetonitrile). Solvent B was used as the mobile phase (0.1% formic acid, 98% acetonitrile). A linear gradient was run to elute the column from 5 to 35% over 60 min at a flow rate of 300 nl/min. Data acquisition was set for one high resolution MS scan in the Orbitrap (60,000 @ 400 m/z). The 6 most abun-

dant ions from the first scan were fragmented by collision- induced dissociation (CID) and the linear trap quadrupole (LTQ) acquired the MS/MS fragment ions. Charge state screening was enabled and unassigned or singly charged ions were rejected. A dynamic exclusion window of 15 s was limited to 300 entries. MS precursors had to exceed 150 ion counts for MS/MS scans. Accumulation time of ions was set to 500 ms (MS 1) and 250 ms (MS 2) using a target setting of 106 (MS 1) and 104 (MS 2). A reference sample containing 200 fmol of human Glu1-fibrinopeptide B (Sigma) was analyzed to monitor the LC- MS/MS systems performance after each batch.

Level 1 data spectra were analyzed with X!Tandem against the canonical human proteome reference dataset, with reverse decoy sequences. The search parameters were set to include only fully tryptic peptides (KR/P) containing up to two missed cleavages. Carbamidomethyl on cysteines was set as a peptide modification. Oxidation on methiones and phosphorylation on serines, threonines, and tyrosines were set as peptide modifications. The precursor mass tolerance was set to 25 ppm, the fragment mass error tolerance to 0.5 Da. Peptides were statistically evaluated using PeptideProphet and protein inference by ProteinProphet, both part of the Trans Proteomic Pipeline. A minimum protein probability of 0.9 was set to match a false discovery rate (FDR) of less than 1%.

Level 2 data comparison of IgG pulldown with each treatment condition was performed using the SAINT method [Choi et al., 2002] using the SAINTexpress software [Teo et al., 2014]. Proteins below a corrected FDR of 10% in any condition were retained and compared against commonly detected proteins in AP-MS datasets (Crapome) [Mellacheruvu et al., 2013]. Databases of previously identified AR interactors were compared with the AP-MS data from VCaP. Path-

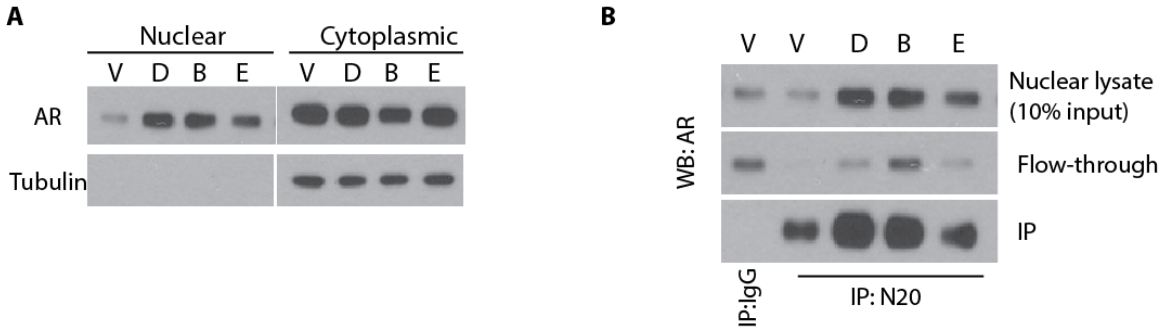

Figure 6.2: **AR antibody pulls down complexes and negative control IgG pull down does not.**
(A) Western blot for AR following drug vehicle (V), ligand DHT (D), agonist bicalutamide (B), and antagonist enzalutamide (E) for nuclear- and cytoplasma-enriched fractions.
(B) ImmunoPrep of AR in the experiment shows reasonable efficiency using the N20 antibody. Control IgG is not pulling down AR.

wayCommons v8 [Cerami et al., 2011], stringDB v10 [Szklarczyk et al., 2015], and McGill AR interactors database [Gottlieb, 1998].

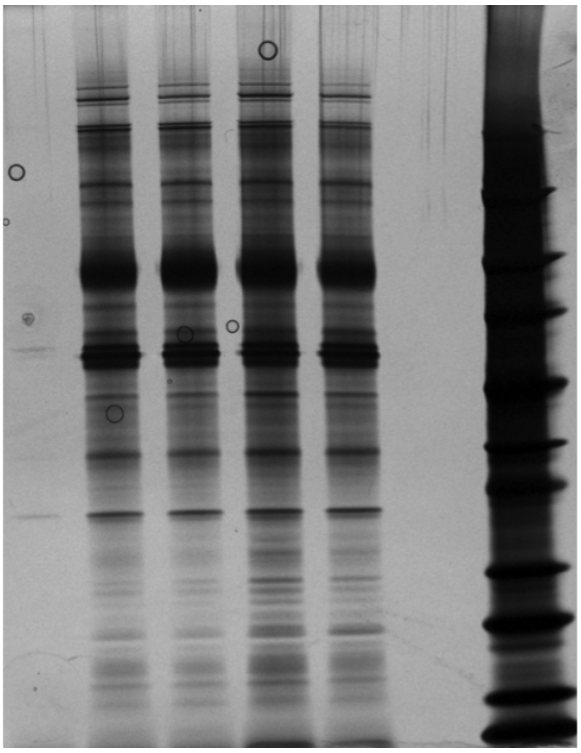

Figure 6.3: **Silver stain of AR pull-downs following perturbations shows numerous interactors but few changes across perturbations.** There are a few observable changes in bands across conditions but in general not large differences among conditions, with the caveat that silver stain is not always a particularly sensitive technique. (V) drug vehicle, (D) agonist DHT, (B) bicalutamide, (M) antagonist MDV3100 or enzalutamide. The negative control lane shows few bands as expected.

6.3 Results - identification of AR interactors and effects of drug perturbations

I used an *in vitro* prostate cancer model of CRPC perturbed with clinically-relevant drugs and identified proteins that interact with AR by AP-MS. Pull down efficiency was assessed initially by western blot shown in figure 6.2, which shows that IgG negative control is not pulling down AR and both enzalutamide and drug vehicle treatment results in lower nuclear AR than DHT and

bicalutamide. A silver stain shows no massive differences among perturbation conditions, shown in figure 6.3. I identified many previously reported AR interactors including PRKDC, ARID1A, ARID1B, SMARCA4, SMARCC1, USP7, and HSP90AA1/AB1. I also identified over 50 putative interactors that are unreported in the PathwayCommons database, including ERF, MTDH, CHTF8, BAG2 and PGAM5.

A total of 111 proteins were quantified across four treatments, with 105 of those proteins present in ligand DHT treatment or vehicle shown in figure 6.4. Gene ontology molecular function categories were mapped to each protein using biomaRt. This resulted in more than 1,000 ontology categories that were manually simplified into eight categories shown in figure 6.4B. Significantly fewer interactors were identified in the cytoplasmic compared to the nuclear fractions, figures 6.4A and 6.5. Approximately one-third of interactors contained AR sequence recognition motifs LXXLL or FXXLF, figures 6.6 and 6.7.

I found minimal differences in AR interactions between agonist and antagonist treatment. Figure 6.8 shows a plot of mean spectral counts for enzalutamide treatment vs DHT treatment. A number of proteins show differences in their means; however, further examination of their variance or behavior in drug vehicle treatment suggests that there is not a statistically significant difference, which was confirmed by unpaired t-test with multiple hypothesis testing correction.

Many interactors are recurrently altered in more than 20% of patient cohorts and may represent promising drug targets, see figure 6.9. Some interactors recurrently altered in patient cohorts above 20%, including CHTF8 which shows a pattern of loss in metastatic samples consistent with its known roles as a tumor

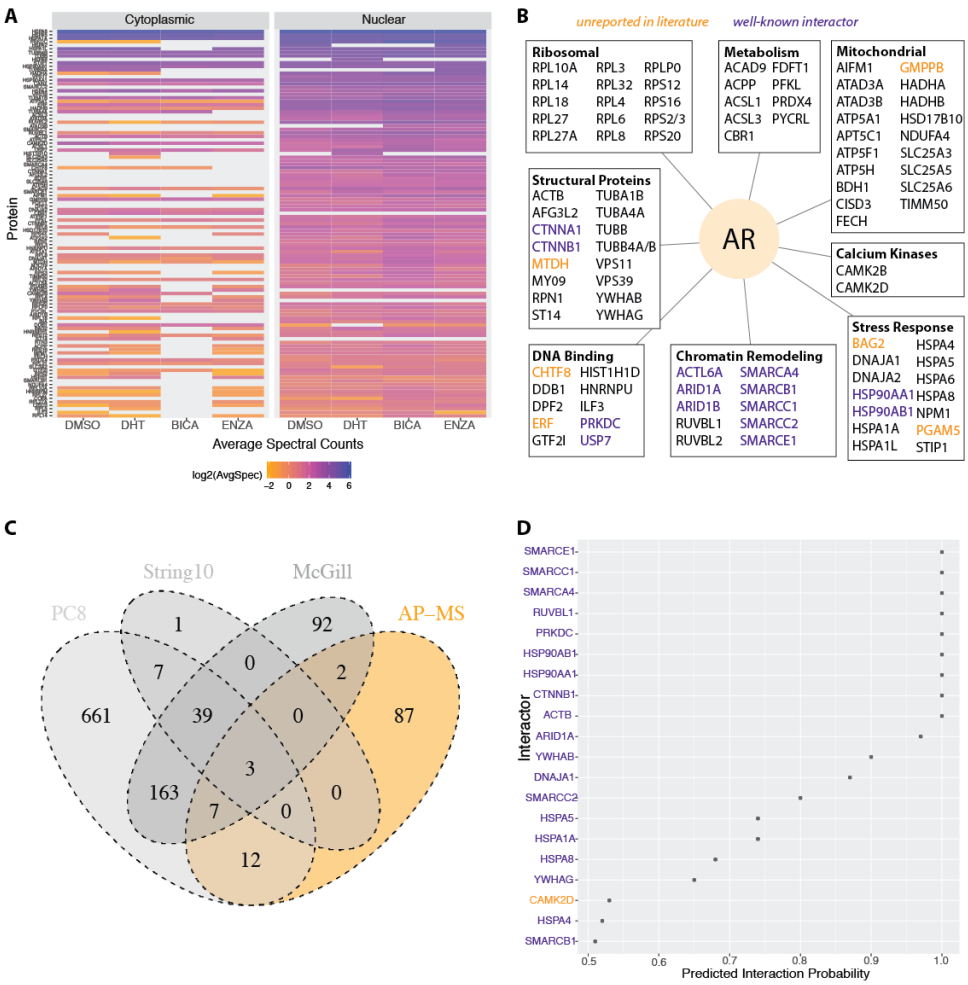

Figure 6.4: **AP-MS identifies numerous, diverse, and novel AR-interacting proteins.**
(A) Log2 of protein spectral counts for proteins with a Bayesian false-discovery-rate less than 0.05 compared to IgG negative control as determined by SAINTexpress. Missing values shown in gray do not necessarily indicate the absence of a proteins expression.
(B) Simplified ontology categories for the AP-MS proteins show a large diversity in functions. GO categories for all the proteins were manually simplified into eight categories.
(C) Overlap of AP-MS with known interactors reveals novel interactors. AR interactors reported in several databases show limited agreement with interactors found by AP-MS in VCaP. PathwayCommons (PC8) is a union of public pathway databases; String (version 10) was limited to high confidence, experimentally determined interactors; McGill is a manually curated database of AR interactors.
(D) Computationally predicted probability of AP-MS interactions determined by PrePPI method, which uses structural, functional, evolutionary, and protein expression (Chen/Honig, PLoS Comp Bio 2015)

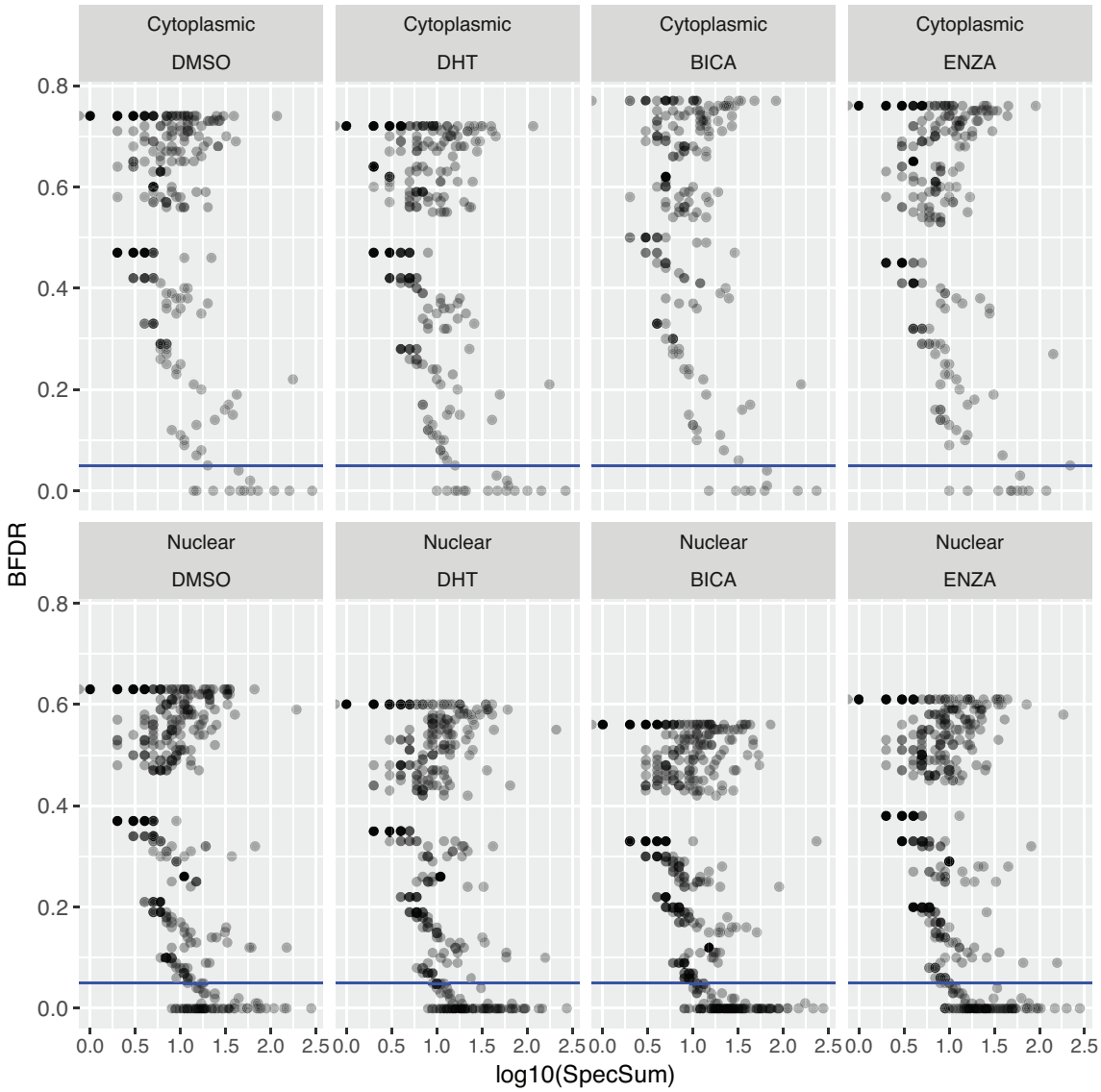

Figure 6.5: **Statistical assessment of differential expression between negative control vs. perturbations.** Peptide-to-protein summarization and quantification of spectral counts is performed with the Transproteomic Pipeline search engines. Statistical assessment comparing each perturbation condition against IgG pulldown was performed with SAINTexpress. It computes a Bayesian false-discovery-rate (BFDR) for each protein, corrected for multiple hypothesis testing. Proteins with high spectral counts are more likely to show statistically significant differences with negative control IgG pull-down.

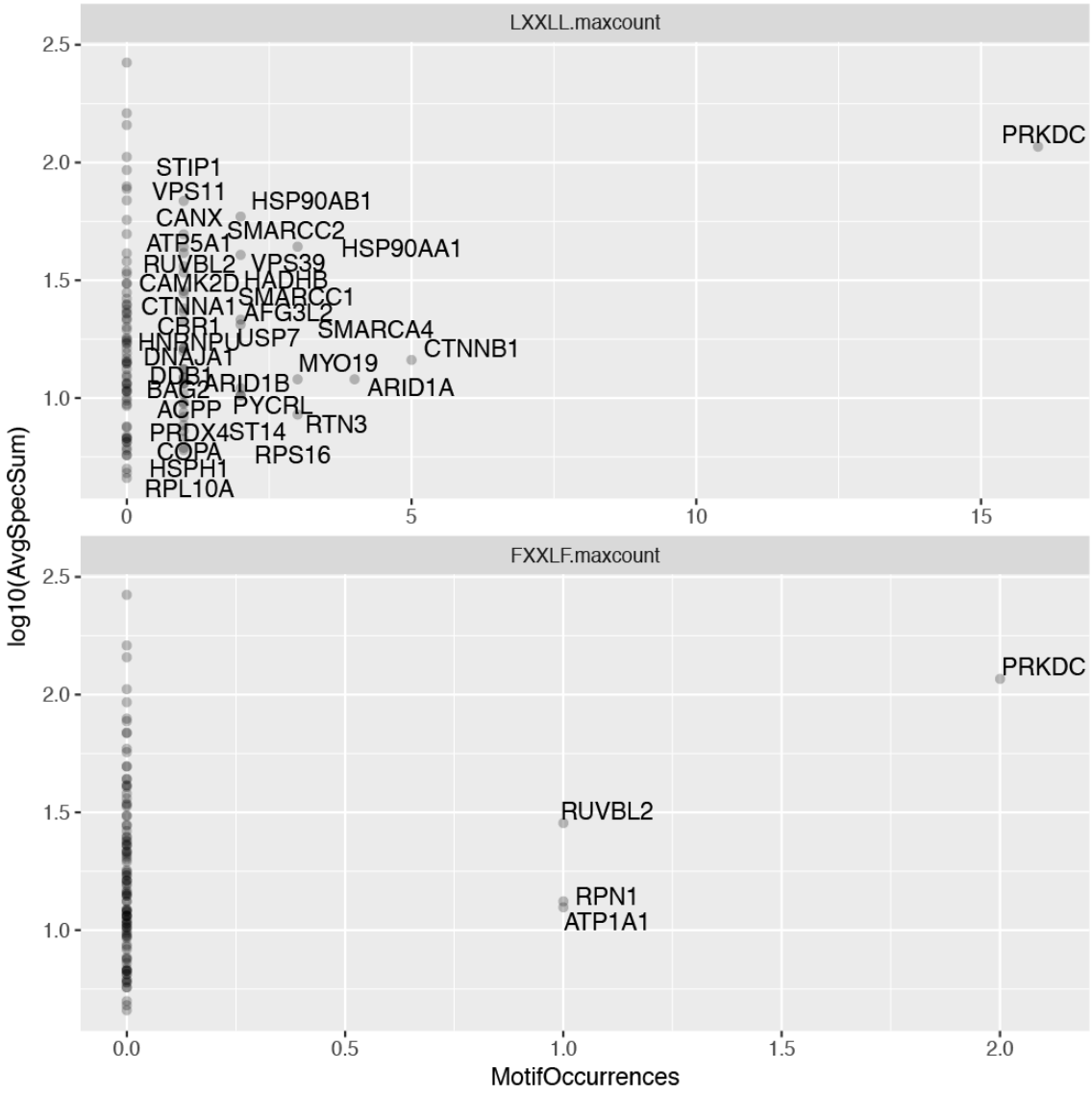

Figure 6.6: **Occurrences of AR-interaction motifs in interactors. Androgen receptor is known to bind to LXXLL and FXXLF motifs.** Approximately 1/3 of the interactors identified by AP-MS contain at least 1 of these motifs. PRKDC is an outlier containing numerous occurrences of the LXXLL motif. Presence of either motif is evidence that an identified interactor is less likely to be a false positive. For example, BAG2 and CAM2KD are not reported as AR interactors in PathwayCommons, McGill AR database, or StringDB but they each contain an LXXLL motif.

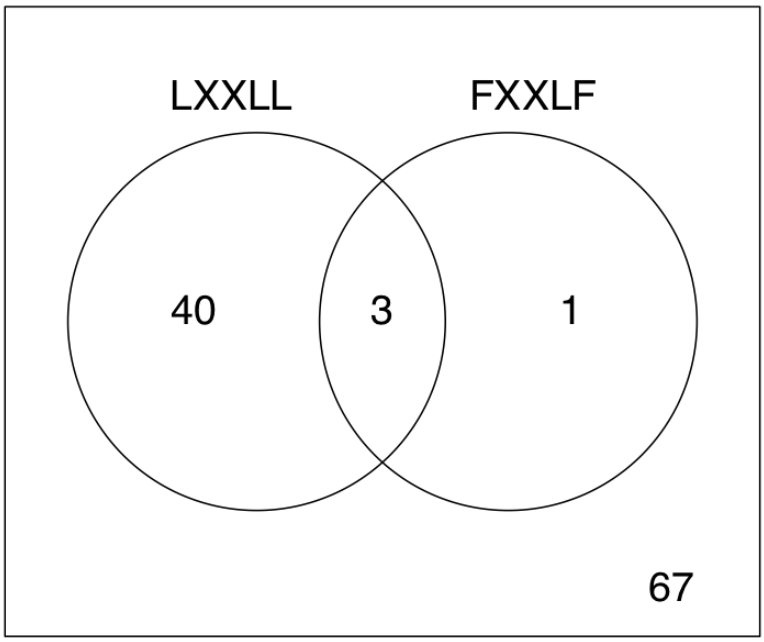

Figure 6.7: **Distribution and overlap of AR interaction motifs among identified interactors.**

suppressor. MTDH shows a pattern of amplification consistent with its known roles in epitheial-to-mesenchymal transition.

6.4 Discussion

Using AP-MS and the VCaP model perturbed with clinically-relevant drugs I identified a large number of AR interacting proteins that are previously unreported and confirmed a number of well-known interactors including SWI/SNF, PRKDC and chaperone proteins. AR interactors appear to be robust against drug treatment with agonist bicalutamide or antagonist enzalutamide, suggesting that interactomes may not be as dynamic as is sometimes described in the literature [Paliouras et al., 2011, Mohammed et al., 2013]. These results also raise the possibility that although enzalutamide blocks some nuclear translocation

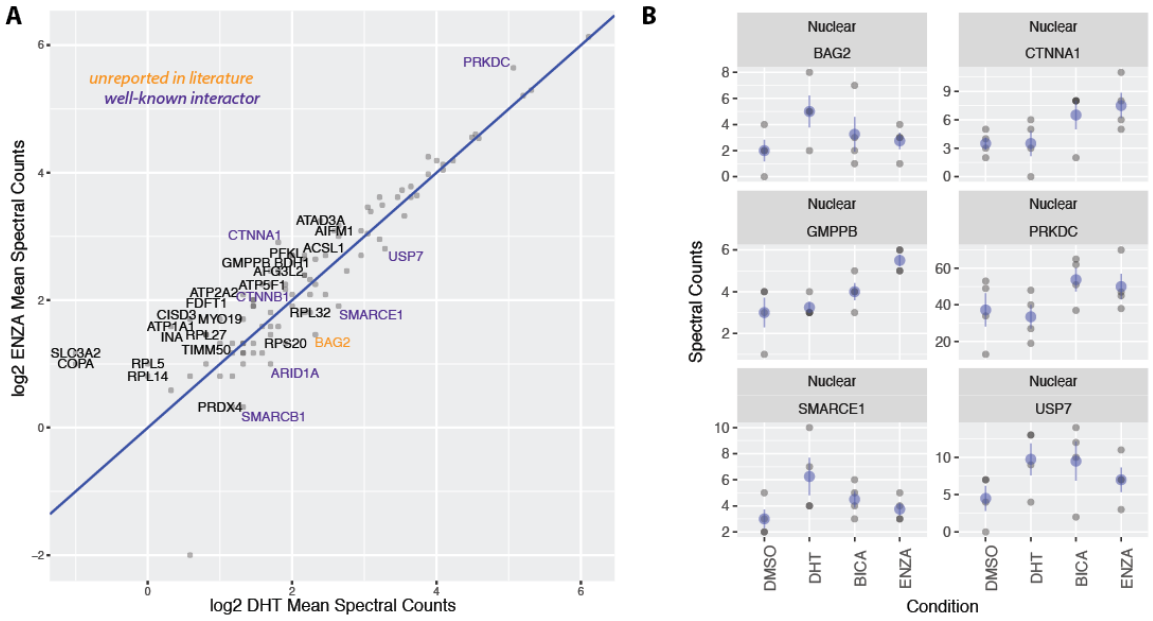

Figure 6.8: **AR-Interacting proteins are present in similar abundances across ligand and antagonist treatments.**
(A) Plot of ligand DHT vs antagonist enzalutamide shows most proteins have similar expression levels in both treatments. Statistical analysis with SAINTexpress indicated no statistically significant differences in expressed proteins across the treatment conditions. SAINTexpress incorporates spectral counts and percent of a proteins amino acid chain detected.
(B) Six proteins are shown for biological interest. Black dots indicate measurements in 4 biological replicates. Blue dots indicate mean values and blue lines indicate +/- 1.5 standard deviations. BAG2 promotes protein release from the chaperone HSP70. CTNNA1 works in cell adhesion by interacting with cadherins. GMPPB is thought to be involved in oligosaccharide metabolism. PRKDC is a serine/threonine kinase that senses DNA damage. SMARCE1 is part of the SWI/SNF complex and is involved in chromatin modification. USP7 is a deubitquitinase.

of AR, this does not result in large effects on interactors [Tran et al., 2009]. It is certainly also possible that there are large effects but the 2 hour time point that I chose is too early. A time-course experiment to investigate this further is warranted. It is also possible that affinity purification and subsequent sample processing lacks the requisite sensitivity to find differences in interacting proteins.

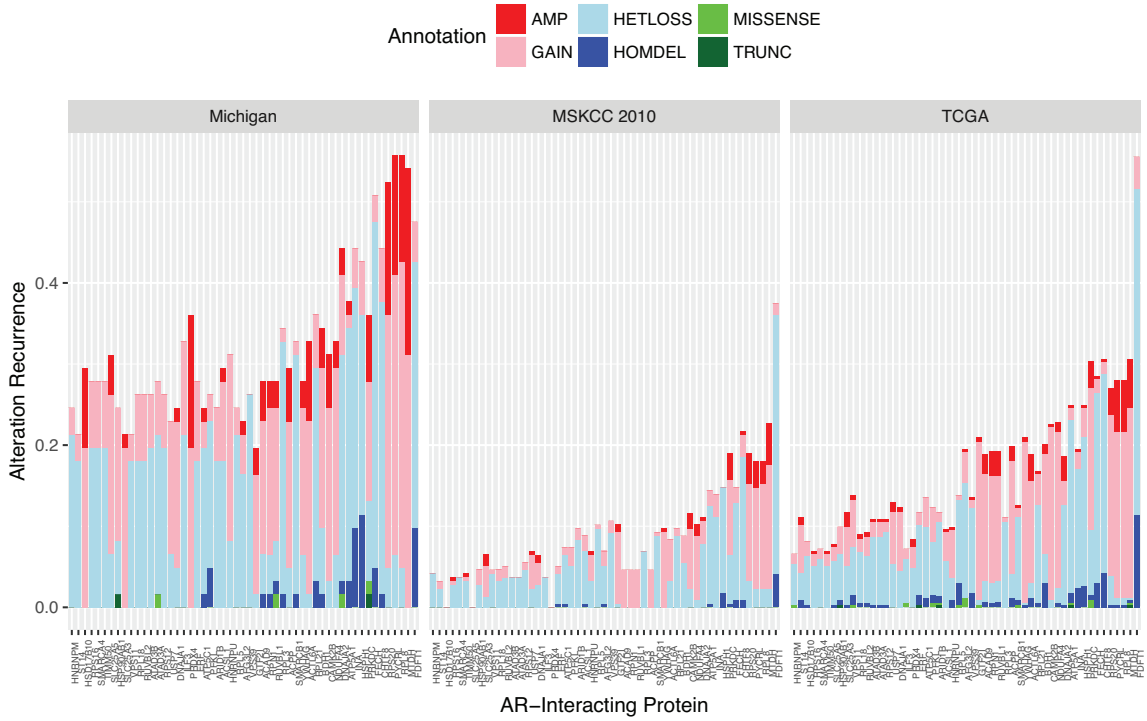

Figure 6.9: **Recurrent alterations in AR-interacting proteins in large clinical cohorts.** Most AR interacting proteins are recurrently altered in less than 20% of patient samples in primary and metastatic prostate cancer at the levels of DNA mutations, copy number deletions/amplifications and mRNA expression high/low. Data for primary prostate tumors comes from a TCGA 2015 study with more than 300 patient samples. Metastatic tumor data in Robinson et al., Cell 2015 (Michigan) has more than 100 patient samples. In metastatic tumors amplifications occur in metadherin MTDH, which is known to activate NF-kappaB. [Created with cBioPortal for cancer genomics [Cerami et al., 2012]]

Techniques such as FLAG-tagging or cross-linking could improve sensitivity. For example, Mohammed and colleagues used cross-linking to examine ligand (estrogen) vs antagonist (tamoxifen) effects on ER-protein interactions and found large differences with GREB1 by SILAC ratio [Mohammed et al., 2013]. I chose a label-free relative quantification strategy. It is possible that I could improve sensitivity and decrease noise using label-based relative quantification, such as SILAC or TMT [Ebhardt et al., 2015].

AR interactors represent a category of potential targets for prostate cancer drug development [Spratt et al., 2016]. Co-activator NCOA2 is recurrently amplified in 6% of CRPC patients and NCOA3 is a target of ubiquitin ligase SPOP [Spratt et al., 2016]. Surpisingly, neither co-factor was identified here by AP-MS, suggesting that the N20 antibody used for affinity purification may have disrupted these interactions with AR. Metadherin (MTDH) is an interactor identified here that is also recurrently amplified in metastatic tumors and is known to activate NF-kappaB and play roles in epithelial-to-mesenchymal transition. Reciprocal pulldown experiments and genetic knockdowns are necessary experiments to demonstrate that it both interacts with AR and that its knockdown has consequences in phenotypic assays. Wei and colleagues demonstrate that knockdown of MTDH effected viability, apoptosis, sensitivity to cisplatin and migration in PC3, LNCaP and DU145 [Wei et al., 2015]. They also found crosstalk with PI3K signaling [Wei et al., 2015].

The AP-MS experiments presented here identified unreported interactors, confirmed known interactors, and suggest that drug perturbations may not have much effect on AR interacting proteins. Metadherin was identified as a previously unreported interactor that is recurrently altered in patient cohorts and warrants further testing. Dynamic measurements of AR interacting proteins using different experimental techniques are also warranted given the importance of AR to prostate tumorigenesis and the possibility of blocking AR-signaling by co-targeting multiple proteins with combination therapy.

CHAPTER 7

CONCLUSION & FUTURE DIRECTIONS FOR PROSTATE CANCER COMBINATION THERAPY

7.1 Targeted proteomics and perturbation biology are a promising approach to develop combination therapy

By using a perturbation biology approach with targeted proteomics and network analysis, I was able to identify (phospho)-protein markers of response and resistance to clinically-relevant drug perturbations after short-term exposure. Proteomics measurements could be improved by incorporating labels for relative quantitation, such as TMT or by using state-of-the-art parallel reaction monitoring (PRM). Relatively simple network analysis led to the predictions of more than 10 novel drug combinations. My hypothesis is that these combinations would either block resistance mechanisms or enhance drug potency. A highly potent drug combination will reduce the cancer population numbers and/or growth rate, thereby significantly lowering the probability of the emergence of resistance.

Ten drugs were tested in all one-, two- and three-drug combinations with AR, AKT, MTOR and taxane therapy using phenotypic assays and discovery proteomics (14-3-3 inhibitor only) resulting in two promising drug combinations that warrant further testing and eight that probably do not. The first strategy is AKT inhibitor MK2206 + CDK1/2/5 inhibitor dinaciclib. This is a somewhat novel result because although cell cycle inhibitors are currently being tested in clinical trials in prostate cancer, most inhibitors target CDK4/6 and

not CDK1/2/5 [Spratt et al., 2016]. I am also confident in this strategy because Hu and colleagues reported drug synergy with this combination and launched an NCICTEP-approved multicenter phase I clinical trial in pancreatic cancer [Hu et al., 2015].

The second combination strategy is AR inhibitor enzalutamide + AKT inhibitor MK2206 + 14-3-3 inhibitor BV02. This strategy shows molecular and phenotypic evidence of efficacy. Moreover, there are recurrent alterations in YWHAZ in patient cohorts with prognostic value. Due to concerns with the specificity and efficacy of BV02, experiments are being planned with siRNA knockdown and will be carried out by H. Alexander Ebhardt's laboratory at University College, Dublin. Further molecular and phenotypic measurements at the preclinical level are warranted before pursuing either strategy in patients. In particular, xenograft or GEM experiments would be helpful to confirm efficacy of strategy I and II *in vivo*. Molecular measurements of (phospho)-protein responses are also warranted including: phospho-RB to assess cell cycle; phospho-AKT, phospho-p70S6K, phospho-4EPB1 to assess PI3K signaling activity; AR target FKBP5 to assess AR activity; cleaved-PARP and cleaved-caspase3/7 to assess apoptosis; ki67 to assess proliferation. If these preclinical data are promising then discussions about whether to pursue clinical trials are warranted.

7.2 A diversity of approaches to develop combination therapy should continue to be pursued

Developing combination therapy is a hard problem, partly because the search space grows exponentially with each additional variable [Bulusu et al., 2016]. However, there are a number of recent advances in pre-clinical and clinical research that give reasons for optimism. Recent work by Zimmer and colleagues found that *in vitro* phenotypic drug effects in higher order drug combinations could be predicted from only measurements of drug pairs [Zimmer et al., 2016]. There is a long and rich tradition in physics and biology of predicting higher order effects from pairwise interactions, including the EVfold method for predicting protein structure from co-evolution of amino-acid sequences [Marks et al., 2011]. These findings suggest that pairwise screening of drug and drug-like libraries in cancer cell line panels may enable successful prediction of n-drug combinations. At the same time, advances in creating prostate organoids and patient-derived xenograft models are enabling drug development, particularly for prostate cancers that are not well-represented by the commonly used cell lines LNCaP, VCaP, PC3, DU145 and 22Rv1 [Pauli et al., 2017].

In addition to advances in pre-clinical research for combination therapy development, there has also been progress in our understanding of the molecular alterations and tumor heterogeneity in prostate cancer patients [Taylor et al., 2010, Chen et al., 2013, Boutros et al., 2015, Robinson et al., 2015, Drake et al., 2016, Spratt et al., 2016, Blattner et al., 2017]. Combination therapy including immuno-therapies may greatly augment drug combinations [Fong et al., 2009]. Efforts to create shared databases that match molecu-

lar alterations and germline DNA with drug sensitivity and toxicity profiles will help direct patients to clinical trials to enable personalized medicine [Cerami et al., 2012, Huang et al., 2017] similar to what was achieved in HIV but at a greater level of complexity [Bock and Lengauer, 2012]. International efforts to coordinate sharing of clinical data will facilitate personalized medicine [Vis et al., 2017, Siu et al., 2016].

Advances in molecular measurement technologies and economies of scale now make it feasible to sequence prostate cancer genomes for both research purposes and integration into clinical support systems [Pauli et al., 2017]. Advances in high resolution chromatography and mass spectrometry proteomics are augmenting genomic information and providing information about cellular signaling networks through measurements of regulatory modifications, such as phosphorylation [Cifani et al., 2017, Mertins et al., 2016]. By integrating the two, proteogenomics provides insights into DNA-to-RNA-to-protein regulation, cellular signaling networks, disease subtypes, and biomarkers for diagnosis or treatment prediction [Ruggles et al., 2017, Nesvizhskii, 2014]. Large-scale pharmaco-genomic studies aim to predict one-, two-, multi-drug sensitivities from tumor genomes [Iorio et al., 2016]. The addition of proteomic measurements will provide information about signaling and metabolic networks that may lead to an emerging field of pharmaco-proteogenomics with even greater ability to predict drug effects.

Early detection and biomarker development efforts, including hereditary testing for DNA repair defects may improve patient stratification and treatment outcomes [Evans et al., 2016]. Although the history of biomarker research is plagued by failures [Ioannidis, 2013], there is new optimism that improvements

in high resolution chromatography, mass spectrometry, circulating tumor DNA, and a greater rigor in statistical methodologies and study designs may finally discover new clinically useful biomarkers [Cifani et al., 2017, Skates et al., 2013]. In conclusion, there is a great diversity of approaches having success in prostate cancer research that have only been briefly described here that are contributing to developing effective combination therapies.

APPENDIX A

APPENDIX - REVIEW ARTICLE REPRINT: APPLICATIONS OF TARGETED PROTEOMICS IN SYSTEMS BIOLOGY AND TRANSLATIONAL MEDICINE

REVIEW

Applications of targeted proteomics in systems biology and translational medicine

H. Alexander Ebhardt[1], Alex Root[2,3], Chris Sander[2] and Ruedi Aebersold[1,4]

[1] Department of Biology, Institute of Molecular Systems Biology, Eidgenossische Technische Hochschule (ETH) Zurich, Zurich, Switzerland
[2] Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
[3] Department of Physiology, Biophysics and Systems Biology, Weill Cornell Medical College, New York, NY, USA
[4] Faculty of Science, University of Zurich, Zurich, Switzerland

Biological systems are composed of numerous components of which proteins are of particularly high functional significance. Network models are useful abstractions for studying these components in context. Network representations display molecules as nodes and their interactions as edges. Because they are difficult to directly measure, functional edges are frequently inferred from suitably structured datasets consisting of the accurate and consistent quantification of network nodes under a multitude of perturbed conditions. For the precise quantification of a finite list of proteins across a wide range of samples, targeted proteomics exemplified by selected/multiple reaction monitoring (SRM, MRM) mass spectrometry has proven useful and has been applied to a variety of questions in systems biology and clinical studies. Here, we survey the literature of studies using SRM-MS in systems biology and clinical proteomics. Systems biology studies frequently examine fundamental questions in network biology, whereas clinical studies frequently focus on biomarker discovery and validation in a variety of diseases including cardiovascular disease and cancer. Targeted proteomics promises to advance our understanding of biological networks and the phenotypic significance of specific network states and to advance biomarkers into clinical use.

Received: January 8, 2015
Revised: April 27, 2015
Accepted: June 9, 2015

Keywords:
Clinical proteomics / Multiple reaction monitoring / Selected reaction monitoring / Systems biology / Targeted proteomics

1 Introduction

1.1 Why networks?

Many observations related to signaling cascades and other biological processes cannot be explained with a simple linear network model. This has led to the realization that the components of biological systems are not simply connected in a linear fashion but through a web of interactions, feedback loops, and crosstalk at multiple spatial and temporal scales and prompted a shift towards considering these processes as dynamic network. Models of such networks aim at increasing our understanding how changes in network state in specific contexts throughout normal development, disease, and in response to perturbations [1–3] generate or modulate phenotypes. A molecular network is an abstract construct in which nodes represent molecules and the edges signify a variety of physical or functional types of interactions. Physical interactions include interactions between proteins, enzyme–substrate relationships, transcription factors and their target genes, and protein–RNA interactions. Functional interactions include transient enzyme–substrate interactions, genetic interactions7&N and other functional dependencies of presently unknown mechanism. Because functional interactions are difficult to directly measure they are frequently

Correspondence: Dr. H. Alexander Ebhardt, ETH – D-BIOL – IMSB, Auguste-Piccard-Hof 1, 8093 Zurich, Switzerland
Email: ebhardt@imsb.biol.ethz.ch
Fax: + 41 44 633 15 32

Abbreviations: **AIMS**, accurate inclusion mass scanning; **CPTAC**, clinical proteomics tumor analysis consortium; **ESI**, electrospray ionization; **HPLC**, high pressure liquid chromatography; **MRM**, multiple reaction monitoring; **MS**, mass spectrometry; **MS/MS**, tandem mass spectrometry; **PTM**, post-translational modification; **RPPA**, reverse-phase protein array; **RT**, retention time; **SID**, stable isotope dilution; **SISCAPA**, stable isotope standards and capture by anti-peptide antibodies; **SRM**, selected reaction monitoring; **QQQ**, triple quadrupole

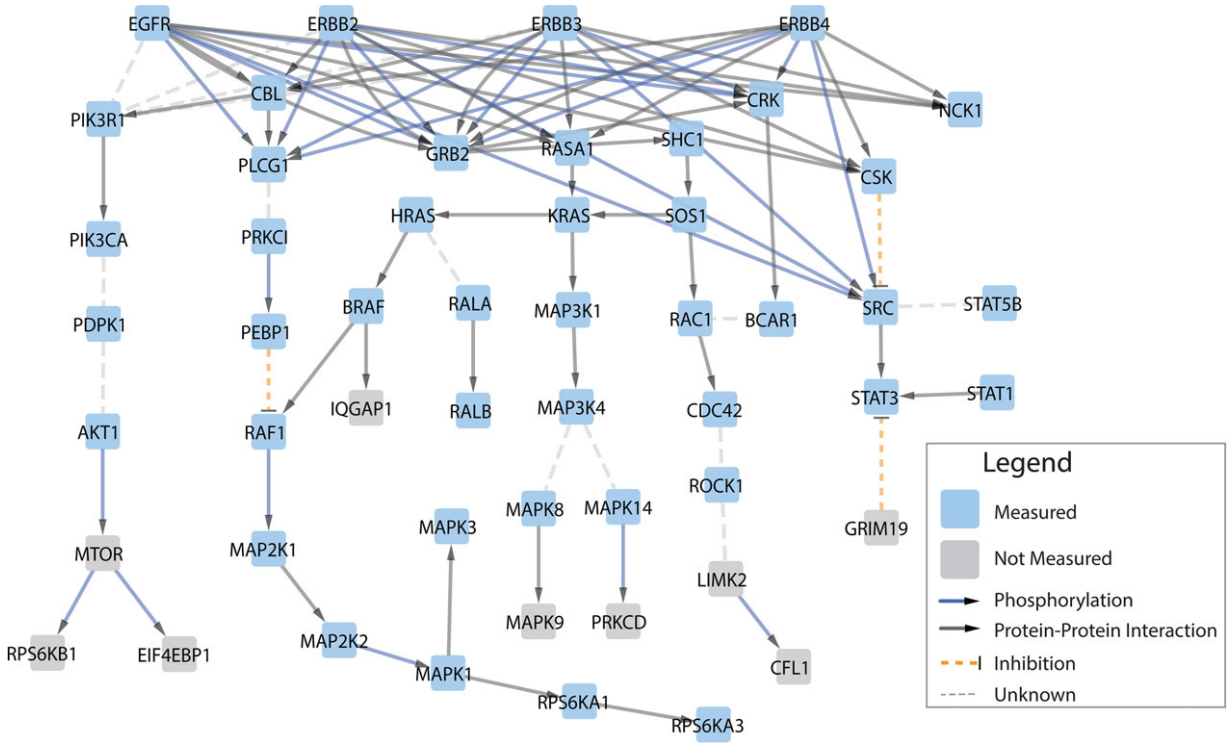

Figure 1. Network biology paradigm and complexities of proteomes. **(A)** Network biology paradigm. Protein–protein interactions can be modeled as networks involving a variety of interaction types. **(B)** A few complexities of the proteome. Studying proteins is complicated due to several factors: (i) a typical cell contains in excess of 20 000 different proteins, isoforms, and post-translational modifications (PTMs); (ii) the range of absolute abundances spans more than seven orders of magnitude; (iii) each cell, tissue, and organism has a different complement of proteins; (iv) proteins vary in space and (v) in time; (vi) proteins are involved in numerous interactions subject to context-dependent "rewiring".

inferred by statistical correlation from suitably structured, large datasets [4–6]. Networks also provide a unifying conceptual and mathematical framework for systems biology, ecology, and neuroscience [7]. An example of a biological network from the recent literature [8] is shown in Fig. 1. Alongside conceptual advances, immense technological progress now supports the identification and quantification of nucleic acids, proteins, lipids, carbohydrates, metabolites, and small molecules at sufficient coverage, depth, and throughput for the study of biological networks in diverse contexts, such as cell lines, body fluids, or tissues from various organisms and across developmental stages or among phenotypic states [9].

1.2 Why proteins?

Proteins are key components of many types of molecular networks, perform most biochemical functions of the cell and are the targets of most current drugs [10]. Although proteins can be reliably identified by the well known discovery proteomics methods at high throughput [11] and their 3D structures can be determined experimentally or computationally [12], there remain numerous unsolved problems relating to their structure and function in the context of network biology. Measuring proteins poses technical challenges, particularly in higher eukaryotes which are made-up of trillions of cells that are categorized, somewhat arbitrarily, into more than 400 different cell types [13]. The two foremost challenges are the sheer number of proteins present in a cell and their vast dynamic range of expression which spans four to five orders of magnitude in prokaryotes, —six to seven orders of magnitude in eukaryotic cells/tissues and 12 orders of magnitude in body fluids [14–16]. Moreover, proteins are subject to more than 200 types of PTMs, which further magnify the number of distinct protein entities in a sample and sample handling chemistries [17,18]. Figure 2 illustrates these challenges. In all, from a single protein coding genomic loci a myriad of protoforms can arise, especially in higher eukaryotes, which can be identified and quantified using high mass accuracy MS.

Protein networks pose a number of data analysis challenges that are rooted in the fact that proteins cannot be represented as simple molecular entities. Rather, they operate in a large number of biological contexts, spatial, and temporal scales and functional states exemplified by protein-specific properties such as reaction mechanisms, substrate/motif binding and complex formation. In addition, properties of protein networks such as information processing, noise, adaptability, robustness, and even seemingly paradoxical

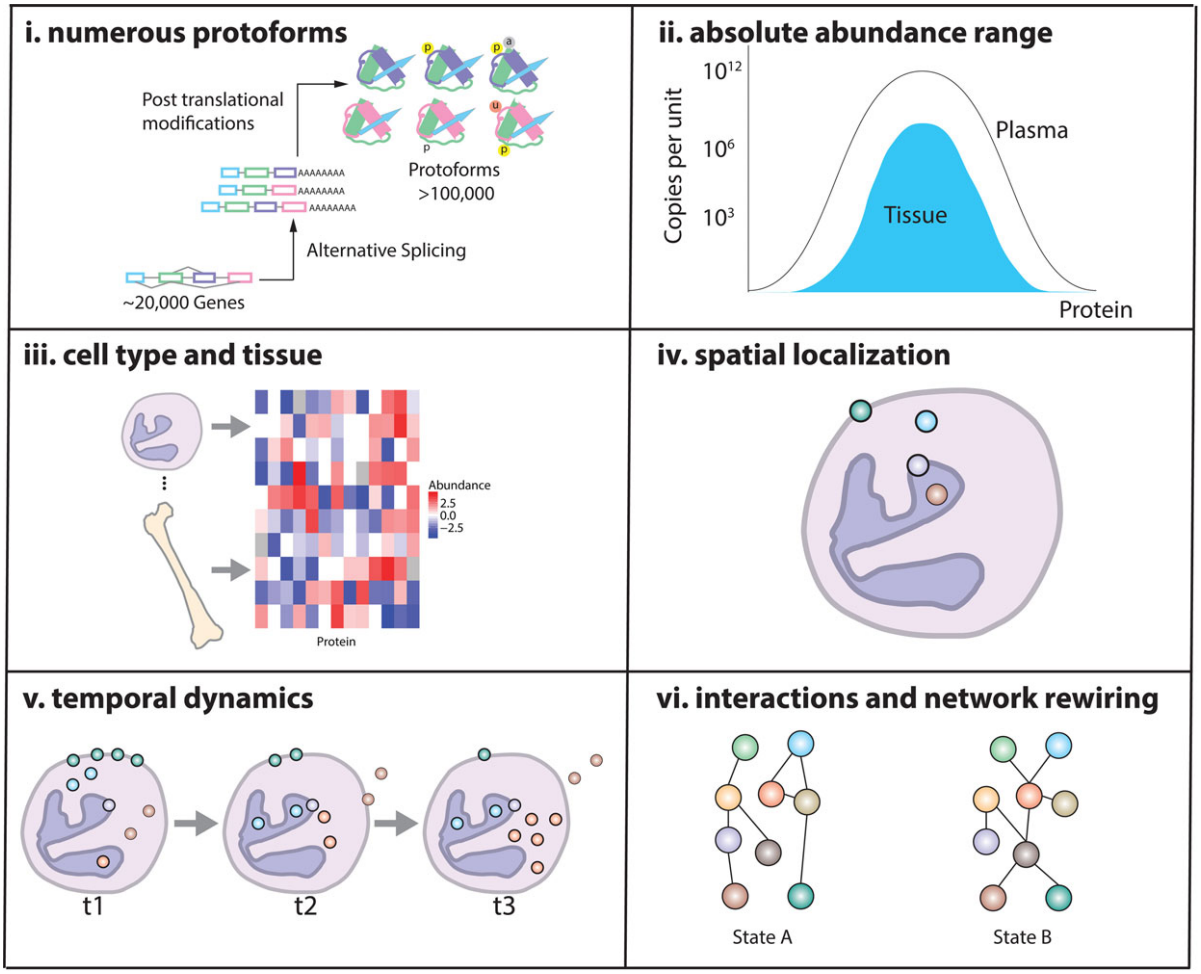

Figure 2. Proteins vary greatly within the cell. There are numerous protoforms to consider which arise from alternative splicing of pre-mRNA and post-tranlational modifications (i). The absolute abundance range of proteins is different in tissue than plasma (ii). Within each cell type, different proteomes are expressed (iii). The spatial localization of proteins also effects the proteins activity (iv). As a function of time and/or stimulus, protein levels and/or spacial distribution might differ (v). The activity of proteins is effected by protein-protein interactions and rewiring of protein networks (vi). All points raised above effect methods to extract the proteome, or parts thereof.

arrangement of components and functions, such as enzyme promiscuity [19–26]. Moreover, a protein can exist with different variant sequences due to splicing or mutations, and be subject to different PTMs at different sites, resulting in a vast number of theoretical combination of PTMs, known as "mod-forms", see for example the histone-code [18]. Even for a protein as well studied as Akt, a recent study found PTMs affecting a new layer of activation mechanism in cell cycle progression [27]. Therefore, while there are compelling reasons to study protein networks of the cell, their analysis challenges current algorithmic and technical capabilities.

1.3 Protein identification and quantification

There are two main ways to detect and quantify proteins: affinity reagent based methods, exemplified by ELISA, Western blotting or immuno histochemistry staining, and MS based peptide identification and quantification, which is mainly used for research and discovery proteomics. However, the dynamic range and number of proteins quantifiable using affinity-reagent based assays is limited [28]. Quantitatively describing protein networks not only over a large dynamic range but also multiple samples in a reproducible manner will lead to a better understanding of the biological protein network and be a better clinical predictor than single protein measurements alone. Targeted proteomics is best suited to meet these needs. Multiple studies clearly demonstrated the reproducibility of SRM-MS across laboratories [29, 30]. SRM-MS has been applied to quantify protein levels of liver tissue across 40 strains of BDX mouse [31] and quantitative trait analysis (QTL) of 78 *Saccharomyces cerevisiae* strains [32]. Also, SRM assays are relatively easy to establish based either on prior knowledge (SRM assay repositories) or rapidly developed using whole protein digest [33]. This is in contrast to establishing a new (batch of) affinity reagent.

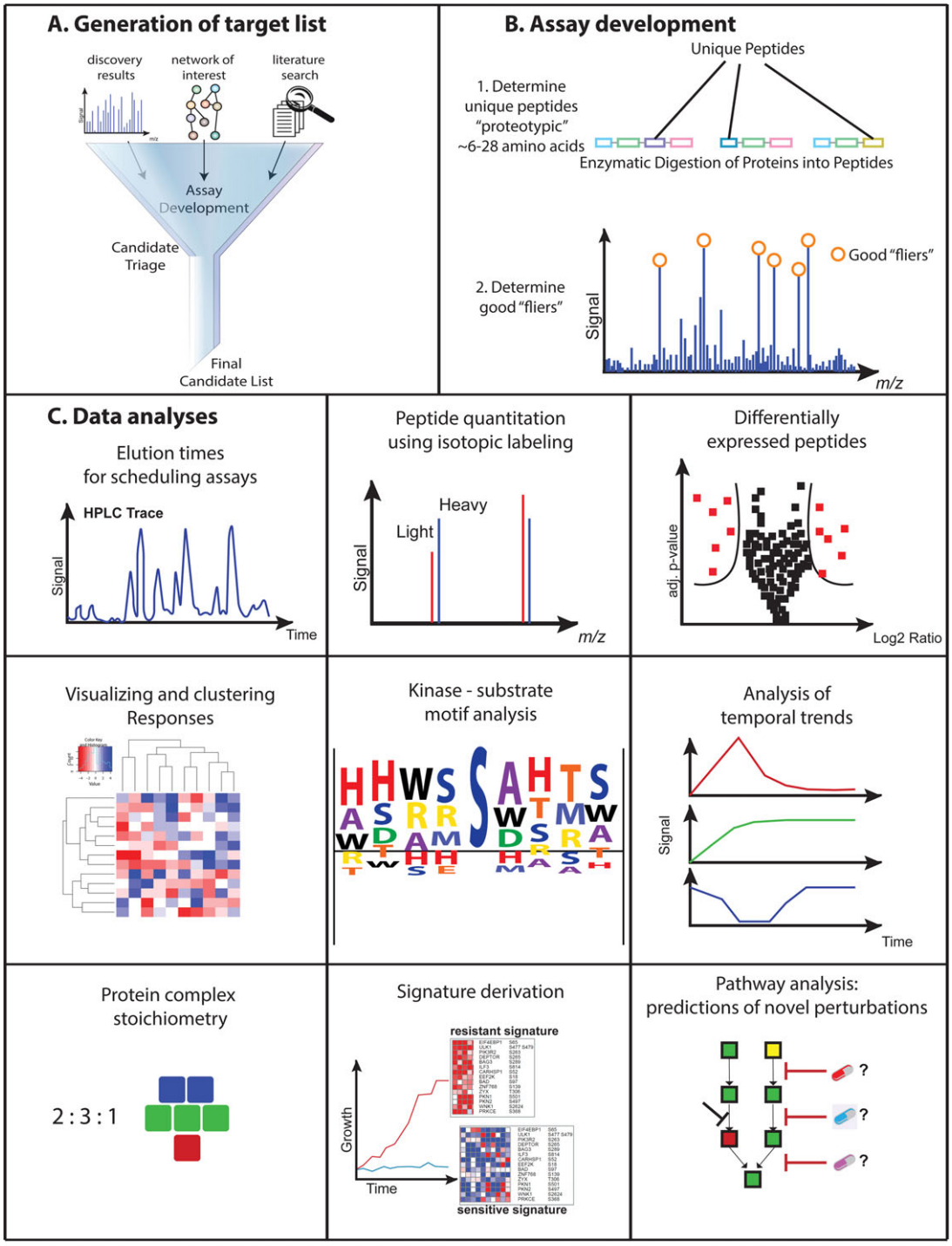

Figure 3. Typical targeted proteomics workflow. A. Discovery results from LC-MS/MS experiments, protein network modeling and literature search typically form the basis to generate the final candidate list to be quantified by SRM. B. SRM assays for peptides are generated from extensive LC-MS/MS experiments under consideration of proteotypic peptides generated and best performing transitions per peptide. C. Data anlysis starts with the primary LC-MS/MS performance examination. If spiked in, stable isotope labeled peptides serve as reference for consistent quantification. Statistical analysis of peptides quantified serve to identify peptides, and therefore proteins, changing in abundance. Further analysis include the clustering of data corresponding to proteins quantified and condition. If multiple kinase substrates were quantified, a consensus motif analysis could identify novel substrate motifs of a kinase. In case the conditions are time course data, the abundance of proteins can be plotted as a function of time. Using SRM-MS, protein stoichiometry of purified protein complexes can be determined (to be precise, this method requires newly synthesized externally calibrated reference peptides). The quantification of proteins and together with sample knowledge integration might lead to signatures which protein signature results in resistant or sensitive samples. The ultimate analysis is the protein network analysis leading to the prediction of novel perturbations.

2 Targeted MS considerations

2.1 Overview: targeted MS workflow

A targeted MS-based proteomics experiment consists of multiple steps and is schematically illustrated in Fig. 3. Specifically, the steps are (1) generation of a hypothesis, a target list of proteins to test the hypothesis and a fit-for-purpose quantitation strategy; (2) study design and experimental planning; (3) sample preparation; (4) method refinement; (5) data acquisition; (6) analysis and modeling [34–36]. Bioinformatics and computational proteomics are part of each step of the workflow. Considerations regarding specificity, precision, and quantitative accuracy effect all steps of the targeted workflow and are discussed in more detail below.

2.2 Step 1: hypothesis development: selection of targets and quantitation strategy

The target list can be composed of several hundred peptides to be measured in a multiplexed fashion, allowing for a wide and versatile set of hypotheses to be tested. Typically the target list will be chosen from biomarker candidates found in a discovery profiling experiment, previous interest in a pathway, or from computational analysis identifying pathways and networks [8]. The total number of analytes quantified per sample injection is typically 50–100 peptides.

In general, to test a hypothesis in a basic science model system, a larger set of proteins is typically measured, while quantifying biomarkers in a clinical setting for making treatment decisions usually involves a smaller set of proteins. The selection of a peptide quantification strategy should follow a fit-for-purpose approach to achieve the right level of specificity, precision, and quantitative accuracy as described recently in a three-tiered system in Carr and colleagues [37]. This system provides clear guidance with respect to the extent of analytical validation required for each major application type, termed Tier 1 to 3 in the publication [37]. In this system, a "labeled internal standard" refers to the use of a consistent spike-in for relative quantitation and most commonly consists of heavy-labeled peptides, while a "reference standard" is as close as possible to the native protein in the sample and subject to all the same sample preparation steps, its absolute abundance is known, and calibration curves of dilutions can be established. The most demanding category are assays for clinical analysis (Tier 1) and require both labeled internal standards and reference standards, and may need to comply with additional regulatory requirements in each country. Common Tier 2 designs include measuring relative changes in protein expression levels and modifications after drug perturbation or disease for non-clinical purposes. Exploratory studies (Tier 3) require some analytical validation, but do not require labeled internal standards or reference standards, and consequently have the lowest assay time development costs. Tier 3 type applications can proceed label-free, at the cost of somewhat reduced quantitative accuracy. Analytical validation or reference standards are typically either chemically synthesized peptides or complex peptide mixtures containing stable isotopes [38].

All other steps of the SRM-MS method development have been described in detail elsewhere [36, 39]. Figure 3 depicts several of these steps and typical downstream analyses. It should be mentioned, that targeted proteomics described here is based on protease digest of whole proteins. The resulting peptides may be proteotypic and uniquely identify a single protein or protein isoform whereas other peptides may be derived from different proteins. In practice, some proteins may only be detected with a single peptide that is shared between closely related proteins, e.g. peptide LVVVGAGGVGK is shared between RASK, RASN, and RASH_HUMAN – rendering the quantification of a specific protein unreliable. A special case of proteotypic peptides are quantotypic peptides which are stoichiometric with total protein abundance and are not influenced by PTMs under the conditions tested [40]. The choice of the quantified peptides and their suitability to serve as surrogates for protein identification and quantification is therefore an important aspect of targeting proteomics measurements.

3 Biology applications

3.1 Protein abundance studies

One of the earliest studies using targeted proteomics absolutely quantified G-coupled receptor rhodopsin using chemically synthesized peptides as calibration standard to quantify endogenous levels of membrane bound rhodopsin [41]. One of the earliest perturbed protein network study was carried out by Picotti and colleagues quantifying proteins of the Krebs Cycle under diauxic shift in *Saccharomyces cerevisiae* [16]. The methodology quickly spread to medium sized target lists and has been used in a variety of basic biology applications, including pharmacology and developmental biology. Zhang and coworkers examined the abundance of two glutathione S-transferase isoforms in human liver cytosol during detoxification [42]. Heikkinen et al. quantified levels of P450, a drug modifying enzyme, in the drug model organism Beagle dog [43]. In developmental biology, Betke et al. examined the differential localization and abundance of G-beta and G-gamma isoforms of G proteins in pre- and post-synaptic fractions isolated from the cortex, cerebellum, hippocampus, and striatum of adult C57Bl6/J mice which showed significant differences in subcellular localization of different isoforms and provided an advance in understanding the roles of various subunits in different brain tissues [44]. Pharmacological and toxicological examinations of 27 cytochrome P450 proteins in Balb/c mouse liver microsomes and tissue lysates from kidney, lung, intestine, heart, and

brain across different developmental stages, including pregnancy were performed by Hersmann and colleagues [45].

Following the successful quantification of single proteins or small protein lists, assays for targeted proteomics were established for entire model organisms [16, 46, 47]. Applications using medium to large-sized target lists as is the case for the study of biological networks have pushed technology development. Chen and coworkers continued the exploratory studies of human liver by targeting 185 proteins previously detected in the Chinese Human Liver Proteome Project and confirmed the presence of 57 targets, 7 of which contained no information in PeptideAtlas, demonstrating the power of community efforts contributing to the completion of the Human Proteome Project [48]. Worboys and coworkers recently developed assays targeting the human kinome, determining both proteotypic and quantotypic peptides for 21% of proteins in the human kinome [40]. Targeted MS has also contributed to an understanding of development in zebrafish. Groh and coworkers used a combined global proteomics and computational approach to generate a candidate list of sex-related development proteins and established roles for ILF2, ILF3, ZGC:195027 and other proteins [49].

Systems investigations of the response to perturbations by targeted proteomics are providing numerous insights in many higher eukaryotes. Zulak and coworkers investigated the effects of methyl jasmonate on terpene synthase enzyme induction and activity in protein extracts of Norway spruce, demonstrating a coordinated network of chemical defense response and a prime example of the robustness of biological systems [50]. Choi and coworkers quantified eight adipokine proteins in response to hydrogen peroxide-induced oxidative stress in adipocytes [51]. Bisson and coworkers used affinity purification and affinity purification MS (AP-MS) to quantify signaling dynamics of 90 proteins in the GRB2 interactome in HEK293T cells after growth-factor stimulation [52]. Xiang and coworkers used cell line models of multiple myeloma to investigate drug resistance of melphalan by comparing signaling, apoptosis-regulating, and DNA repair component proteins, finding a nuclear factor-kappaB signature [53].

Targeted proteomics is increasingly being used to quantify large numbers of proteins and addressing fundamental questions in biological networks. Sabido and coworkers measured 144 proteins in C57BL/6J and 129Sv mice subjected to various periods of high fat diet, revealing activation of either the peroxisomal beta-oxidation pathway or the lipogenesis pathway in each strain, respectively [54]. Kiel and coworkers quantified and localized 75% of an Erbb network of 198 signaling proteins across HEK293, MCF-7, and keratinocytes, determining key quantitative parameters for cell-type-specific computational modeling in this fundamentally important network in cancer biology [8]. The resulting protein signaling network is a complex, yet typical, network of signal transduction governed by protein abundance and protein-protein interactions to convey phosphorylation signaling as seen in Fig. 1.

Targeted MS is increasingly being used to determine the stoichiometries of protein complexes a topic that is of similar importance to the understanding of biological systems through quantitative modeling. For example, monitoring the functional assembly of the human spliceosomal hPrp19/CDC5L complex under various conditions [55], the F_1F_0-ATP synthase super-assembly in H9c2 cardiomyoblasts undergoing cardiac-like differentiation [56] or the determination of context-dependent stoichiometry of the nuclear pore complex in various human cell lines, which showed unanticipated variability [57, 58].

Absolute quantification of proteins using targeted proteomics has considerably matured from single membrane protein quantification to functional stoichiometry determinations of protein complexes and quantification of protein networks. A pioneering approach was recently presented by Soste and coworkers: through literature search and computational prediction methods sentinel proteins were identified which report on the state of signaling pathways in a single SRM-MS analysis. For *Saccharomyces cerevisiae* 157 proteins and 152 phosphorylated peptides were identified to reflect the status of the cellular signaling activity in a single analysis step, thus providing a broad overview of the state of numerous functional networks of the cell [59].

3.2 Post-translational modification studies

Targeted MS has proven invaluable for the study of PTMs. Glinksi and coworkers examined multisite phosphorylation of trehalose-6-phosphate synthase isozymes in vitro in Arabidopsis [60]. The importance of multisite protein phosphorylation and the value of quantifying it by targeted MS were recently shown for the connexin family of proteins and are described in a review by Chen and coworkers [61]. Danielson and coworkers used antibodies specific for 3-nitrotyrosine to quantify the levels of this modification in alpha-synuclein residues [62]. Held and coworkers developed a new method for studying oxidation in response to reactive oxygen species, termed oxMRM [63]. In a tour-de-force application of their method, they examined site-specific cysteine oxidation status of endogenous p53, finding that residue C182 at the dimerization interface of the DNA-binding domain is susceptible to diamide oxidation. Huang and coworkers studied the effects of K63 polyubiquitination of EGFR on its endocytosis and post-endocytotic sorting as mediated by ubiquitin adaptors [64]. Darwanto and coworkers quantified H2B ubiquitination and H3 K79 methylation in the U937 human leukemia cell line and proposed a crosstalk regulatory mechanism between these two modifications [65]. Wolf-Yadlin and coworkers examined an EFGR network of 222 tyrosine phosphopeptides across seven time points following EGF stimulation of 184A1 HMEC cells, demonstrating excellent sensitivity, robust quantitation, and throughput [66]. A useful case study and tutorial for targeted proteomics with enrichment is presented in Rardin and coworkers, in which they detail a method for measuring lysine acetylated peptides from mitochondria

in mouse liver and targeted quantitation of a lysine acetylation site in succinate dehydrogenase A [67].

4 Clinical applications

4.1 Introduction to clinical applications

Clinical applications aim to translate new discoveries and technologies into improving patient outcomes [68] or to increase the understanding of biochemical processes underlying disease etiology. For several decades MS has played a key role in clinical chemistry, particularly for quantifying metabolites and hormones [69, 70]. More recently, much work has gone into the development of targeted proteomics methods supporting clinical studies. Specific issues addressed include handling complexities of tissues and bodily fluids; agreement of community standards; formation of multi-disciplinary research teams, and consortia, such as the Clinical Proteomics Tumor Analysis Consortium (CPTAC); demonstration of inter- and intra-laboratory reproducibility; and proof-of-concept studies in a variety of clinical applications showing the feasibility of generating new clinical tests from biomarker discovery.

4.2 Biomarker development pipeline

The full pipeline of translating biomarker discovery studies into clinical tests with demonstrated health benefits is a multi-year process that is costly, uncertain, and arduous and, because researchers and projects can easily get stuck in the middle of biomarker development projects, has been likened to a tar pit [71, 72]. Moreover, there are a limited number of successful biomarkers approved across all technology platforms, and considerable controversy even surrounds the success stories, such as PSA testing [73]. Although not the focus of this review, it is crucial to note that applied statistical testing methodology for biomarker development has itself advanced alongside clinical chemistry [73]. A conceptual guide to overcoming biomarker challenges is presented by Rifai and coworkers in detail. They define: (i) candidate discovery; (ii) qualification; (iii) verification; (iv) research assay optimization; (v) biomarker validation; (vi) commercialization [72]. As mentioned above, peptide quantification strategy should follow a fit-for-purpose approach to achieve the right level of specificity, precision, and quantitative accuracy as described in a three-tiered system in Carr and colleagues [37]. Further, findings of promising biomarkers or signatures from preclinical studies should be followed by clinical testing. The bulk of recent work has focused on candidate discovery and overcoming the associated challenges related to tissues and bodily fluids. Targeted proteomics can be used to validate biomarkers found in a project's discovery phase across many patient samples with high accuracy and reproducibility [29, 74]. Cima and coworkers successfully applied this strategy by initially using shotgun data to establish a protein list of 44 candidates. Consistent quantification of these 44 proteins across a large patient cohort allowed the establishment of four N-glycosylated protein makers which differentiate patients with a Gleasson score above or below 7 from blood serum [75].

4.3 Community efforts and the clinical proteomics technology assessment for cancer

The complexity in tissues and body fluids presents an enormous signal-to-noise ratio problem [34]. The Human Plasma Proteome Peptide Atlas was launched to provide a knowledgebase for targeted assays [15, 76]. Inspiration for the use of these assays in the clinical setting comes from the small molecule clinical chemistry community where targeted MS is the gold standard for clinical assays quantifying inborn errors of metabolism, drugs and their metabolites, and steroids and biogenic amines [69]. In 2006 the National Cancer Institute (USA) started the Clinical Proteomic Technology Assessment for Cancer (CPTAC) with the aims of evaluating targeted and discovery technologies for quantitative analysis in tissues and biofluids. This program was renewed in 2011 as the CPTAC, which began focusing on applications [77]. CPTAC member laboratories applied standardized methods of multiple reaction monitoring (MRM) and demonstrated reproducibility, precision, sensitive quantitation in tissues and biofluids [78]. Cox and coworkers performed a similar interlaboratory precision study of IGF-1 in plasma across 130 healthy human samples and 22 samples from patients with acromegaly, finding excellent reproducibility [79]. In summary, targeted proteomics has already established its cross-laboratory reproducibility and robustness and will play a major role in validating protein biomarker across large patient cohorts.

4.4 Technology development and enrichment strategies

Over the last decade advances in targeted proteomics have lead to assay sensitivity for plasma proteins in the low ng/ml range. In 2004, Kuhn and coworkers used a depletion of abundant proteins strategy and size exclusion fractionation to quantitate C-reactive protein, a diagnostic marker for rheumatoid arthritis [80]. Anderson and Hunter targeted 53 medium and high abundant proteins in plasma, leading to assays for 47 of the proteins covering 4.5 orders of magnitude with minimal sample preparation [81]. Keshishian and coworkers developed multiplexed assays for six plasma proteins present in 1–10 ng/ml using strong cation exchange chromatography and major abundant protein depletion, but without immunoaffininty enrichment, demonstrating that the abundance range of typical candidate biomarkers (ng/ml) is achievable with targeted proteomics [82]. Fortin and coworkers also achieved ng/ml sensitivity without immunoaffinity enrichment by quantitating prostate-specific antigen (PSA) in sera

from patients with prostate cancer or benign hyperplasia; moreover, their MS results agreed with established ELISA tests for PSA [83]. Shi and coworkers used an antibody-free approach termed PRISM (high-pressure, high-resolution separations coupled with intelligent selection and multiplexing) in order to quantitate PSA levels in the range of 50–100 pg/ml, also with excellent correlation to clinical immunoassays [84]. Shi and coworkers applied the PRISM targeted proteomics to quantitate AGR2 in human urine at serum at concentrations of approximately 130 pg/ml and 10 pg per 100 ug of total protein mass in urine, respectively, and found in a proof-of-concept study of 37 urine samples that AGR2/PSA concentration ratios can distinguish noncancer and cancer [85]. Fallon and coworkers developed assays for 14 UGT1As and UGT2Bs across 60 human liver microsomes and matching S9 samples to evaluate metabolism in drug development [86].

Technology development has also advanced in labeling incorporation technology and throughput. Zhao and coworkers developed a synthetic peptide strategy using ^{18}O labeling strategy for SID-MRM-MS with the ability to produce synthetic peptides for use as internal quantitation standards in only 1 h with excellent stability [87]. They then utilized these labeled peptides for absolute quantitation of candidate hepatocellular carcinoma biomarkers vitronectin and clusterin in undepleted serum samples. Martinez-Morillo and coworkers developed assays for absolute quantification of apolipoprotein E isoforms in cerebrospinal fluid and plasma, which are important in lipid metabolism in the central nervous system and are associated with coronary atherosclerosis and Alzheimer's disease, and also assessed the effects of chemical modifications on selected target peptides' quantitation [88]. Tang and coworkers used major protein depletion and 1D gel separation, starting with less than 100 μl of serum, obtaining reproducible quantitation without internal standards down to 200 pg/ml in assays for PRDX6, ADAM12, PAEP, CGB, and CTSD, which demonstrates that their GeLC-MRM workflow has sufficient throughput, sensitivity, and costs for an initial screening of large numbers of candidate biomarkers [89].

Technology development has gone into using PTMs as handle to enrich for partially low abundant proteins. One such enrichment strategy is the enrichment of N-glycosylated peptides and the subsequent PNGase F-catalyzed conversion of Asn to Asp using solid state extraction method [90, 91] a method which was successfully applied to serum samples [75, 92]. Upon purification, only the modified peptide is quantified. Stahl-Zeng and coworkers applied minimal fractionation of isolated N-glycosites to quantitate plasma proteins over five orders of magnitude, reaching sub-ng/ml range [93]. Zawadzka and coworkers used both targeted and discovery proteomics to quantify a set of approximately 60 phosphopeptides from healthy human plasma following offline chromatography and immobilized metal ion affinity chromatography for phosphopeptide enrichment [94]. Further advances in targeted proteomics of N-glycosites will undoubtedly bring the N-glyco Atlas providing SRM assays for 5568 N-to-D-modified peptide sequences [95]. These SRM-assays were applied to prostate cancer tissue samples to determine aggressiveness of tumors using targeted extraction of peptide sequences from SWATH-MS maps [96].

Another strategy is the enrichment microparticles [97, 98] or exosomes (or extracellular vesicles) articles from bodily fluids, e.g. human urin or serum, followed by protein isolation and quantification of peptides using SRM. The exosome enrichment strategy was applied to various diseases ranging from bladder cancer [99] over diabetic nephropathy [100] to detecting *Mycobacterium tuberculosis* peptides in serum of patients with active or latent *M. tuberculosis* infection [101].

4.5 Immuno-affinity SRM

While immuno-reagents are commonly used to deplete the most abundant proteins in plasma they can also be used for enrichment of low abundant proteins. A commonly used technique is known as stable isotope standards and capture by anti-peptide antibodies (SISCAPA) [102]. Dupuis and coworkers quantified staphylococcal enterotoxins in foods by applying a combination immunocapture and protein standard absolute quantification (PSAQ) method, which uses isotope-labeled enterotoxins as internal standards [103]. Oe and coworkers used immune capture of amyloid betas in cerebral spinal fluid as potential biomarkers of Alzheimer's disease achieving limits of quantitation down to 200 pg/ml [104]. Berna and coworkers used immune capture to develop assays to quantify myosin light chain 1 in rat serum as a biomarker of cardiac necrosis to predict drug-induced cardiotoxicity over a range of 0.13 – 6.62 nM [105]. Nicol and coworkers used an immunoaffinity approach to quantify carincoembryonic antigen (CEA), secretory leukocyte peptidase inhibitor, tissue factor pathway inhibitor 1,2 (TFPI/TFPI2), and metalloproteinase inhibitor 1 (TIMP1) in sera samples from lung cancer patients down to low ng/ml levels [106]. Hoofnagle and coworkers quantified the cancer marker thyroglobulin in serum using an immunoaffinity approach down to a limit of detection of 2.6 ng/ml [107].

Immunoaffinity targeted proteomics can also be used in combination with additional enrichment techniques to reach low limits of quantitation. Ahn and coworkers used a variation of SISCAPA with a combination of phytohemagglutinin-L_4 (L-PHA) for N-linked glycan capture and a monoclonal anti-peptide TIMP1 antibody conjugated to magnetic beads to quantitate the cancer candidate biomarker TIMP1, which is present at approximately 0.8 ng/ml in serum using only 1.7 μl of serum from a patient with colorectal cancer [108]. Using online chromatography for affinity capture instead of offline magnetic beads, Neubert and coworkers quantified a marker for gastroesophageal reflux disease, pepsin/pepsinogen, in the saliva of healthy volunteers down to a range of 0.17 to 0.67 ng/ml, providing the most sensitive and specific test to date [109]. A few years later, Neubert and coworkers made use of robotic sample preparation and sequential immunoaffinity

capture of the protein NGF using magnetic beads followed by online affinity capture of a signature NGF peptide, which is the first report of this combination technique, and allowed for quantitation of NGF to 7.03 to 450 pg/ml in a clinical trial for chronic pain [110].

Immunoaffinity techniques can also be useful for quantitating proteins in the presence of autoantibodies. Thyroglobulin (Tg) is used to monitor patients after treatment for differentiated thyroid carcinoma and is often accompanied by the presence of auto-antibodies that interfere with immunoassays. Kushnir and coworkers overcame these challenges by enriching serum samples with a rabbit polyclonal antibody for Tg and immunoaffinity purification of a signature Tg peptide, with a lower limit of quantitation of 0.5 ng/ml [111].

Immunoaffinity techniques are also amenable to multiplexing. Using an approach known as mass spectrometric immunoassay (MSIA), Krastins and coworkers rapidly developed assays for 16 different target proteins and their isoforms across seven different clinically important areas and ranging in concentration from pg/ml to ng/ml in bona fide clinical samples [112]. Recently, Peterman and coworkers applied the MSIA approach to detect insulin and its analogues using a pan-insulin antibody over a range from 1.5 to 960 pM [113]. Other immune-MRM efforts are conducted by the Paulovich laboratory to overcome limit of detection issues in complex samples and eliminate the current bottleneck of translating biomarkers found in basic science studies to clinical practice [114]. Further advances in the field concern the generation of immuno-MRM monoclonal antibodies suitable for SRM and conventional antibody applications [115].

4.6 Cardiovascular disease applications

Perhaps the main advantage of targeted MS-based proteomics over ELISA assays is their multiplexing capability, which is of key importance for biomarker development in disease applications with many putative biomarkers such as cardiovascular disease and cancer. Kuzyk and coworkers developed a mixture of 45 peptide standards in EDTA-plasma without affinity depletion or enrichment and found that 31 of the 45 are putative markers of cardiovascular disease [116]. Keshishian and coworkers developed quantitative assays without immunoaffinity enrichment for six proteins of clinical relevance to cardiac injury ranging from 2 to 15 ng/ml and measured these proteins across three time points in six patients undergoing alcohol septal ablation for hypertrophic obstructive cardiomyopathy [117]. Addona and coworkers used a combination of discovery and targeted proteomics in the context of planned myocardial infarction (PMI) for treatment of hypertrophic cardiomyopathy and myocardial infarction (MI) [118]. Samples of blood directly from patient hearts before, during, and after PMI allowed identification of 121 candidate biomarker proteins, over 100 of which were novel. Targeted proteomics was then applied to peripheral plasma from controls and patients with PMI or MI, suggesting verification of candidate biomarkers [118]. Huillet and coworkers used the PSAQ-SRM approach to quantitate cardiovascular disease biomarkers LDH-B, CKMB, myoglobin, and troponin I, in serum samples from myocardial infarction patients [119]. Domanski and coworkers developed assays with 135 stable-isotope labeled peptides for quantitation of 67 candidate biomarkers of cardiovascular disease spanning the top seven most abundant orders of magnitude of concentration in whole plasma with a 30 min assay time, performing 85 technical replicates, which showed excellent sensitivity and retention time accuracy [120]. Further, solid state N-glyco enrichment strategies can be used to enrich to monitor cardiac resynchronization therapy in serum from a canine model [121].

4.7 Cancer applications

The heterogeneity and complexity present in the numerous types of cancer presents an enormous opportunity but significant challenges for targeted proteomics to make transformative contributions. Much technology development has gone into tissue preparation and quantitation of candidate biomarkers addressing the dual challenges of using a limited amount of material in which the candidate biomarker may also be low abundant. DeSouza and coworkers quantified a marker of endometrial cancer, pyruvate kinase M1/M2, in biopsied tissue at 85 nmol/g compared to 21–26 nmol/g in nonmalignant tissue using the mTRAQ labeling method, compared to only a 2x elevation initially determined by discovery iTRAQ proteomics, suggesting that the dynamic range for quantitation may be compressed in discovery scans [122]. Chen and coworkers developed quantitative assays for 22 proteins in the Wnt/beta-catenin signaling pathway in colon cancer cell lines and applied them to frozen colon tissue sections and laser capture microdissected tumor cells [123]. Elschenbroich and coworkers combined discovery and targeted proteomics to develop assays for serous-type epithelial ovarian cancer discovering a panel of 51 candidate proteins and then using synthetic peptides (13 proteins) and stable isotope label standards (four proteins) for targeted quantification in ascites in serum, providing proof-of-concept validation for this strategy [124]. Remily-Wood and coworkers used pathway analysis to develop 95 quantitative assays including synthetic peptide standards for proteins of interest in colon, lung, melanoma, leukemias, and myelomas, which are published online in a Quantitative Assay Database [125]. Selevsek and coworkers developed multiplexed assays with stable isotope dilution standards to analyze 16 proteins associated with bladder cancer in urine with excellent analytical performance and limits of quantitation limits in the low ng/ml range [126].

Clinical proteomics in cancer has also addressed challenges of integrating genomic data such as mutations. Wang and coworkers demonstrated that altered protein products resulting from somatic mutations can be quantified by targeted proteomics by developing assays for Ras proteins and

applying them in colorectal and pancreatic tumor tissue and premalignant pancreatic cyst fluids [127]. He and coworkers recently developed assays to quantitate ERG isoforms in TMPRSS2-ERG positive VCaP cell line and two prostate cancer tissue samples [128].

In 2011 a complete verification pipeline of biomarker candidates in plasma was presented in a tour-de-force study by Whiteaker and coworkers [129]. They integrated 13 datasets from discovery proteomics and genomics to arrive at >1000 candidate proteins in a mouse model of breast cancer and used data-dependent prioritization to triage candidates, developing assays for 88 proteins evaluated across 80 plasma samples, finding 36 over-expressed proteins with excellent analytical performance. Their data-dependent triage of candidates used a MS approach termed accurate inclusion mass screening (AIMS), which is essentially an efficient bridge from discovery to targeted proteomics that uses inclusion-list dependent acquisition on an orbitrap mass spectrometer to verify the presence of a candidate (Jaffe et al., 2008).

Hüttenhain and coworkers also completed a tour-de-force complete verification pipeline of biomarker candidates by generating assays for 1000 cancer-related proteins and used a data-dependent triage strategy by first examining candidate proteins' detectability in plasma and urine samples [130]. They subsequently detected 182 proteins in depleted plasma, spanning five orders of magnitude in abundance and 408 proteins in urine. They then profiled 83 patient plasma samples for 34 of the candidate biomarkers using heavy-labeled synthetic peptides, finding that their targeted proteomics assays allowed for reproducible quantitation.

Biomarker discovery and verification has also been pursued for post-translationally modified proteins. Cima and coworkers followed a two-stage strategy for biomarker discovery, starting with a discovery scan comparing the serum N-linked glycoproteome of PTEN conditional knockout model of prostate cancer to wild-type and then developed targeted assays for 39 human orthologs, which were then quantified in the sera of 143 prostate cancer patients and controls over an abundance range of six orders of magnitude [75]. Computational analysis derived a signature for diagnosis and prognosis of prostate cancer. In a followup study, Kalin and coworkers used the N-linked glycoprotein capture and assays to quantitate candidate biomarkers in sera of 57 patients with metastatic prostate cancer [131]. Computational analysis integrated known prognostic factors with candidate N-linked glycoproteins derived new nomograms with potentially improved accuracy. Cerciello and coworkers used a three stage approach for identification of biomarker candidates in the serum of malignant pleural mesothelioma [132]. First they screened a collection of relevant cell lines and discovered 125 candidate cell surface N-linked glycoprotein peptides. They then developed assays for 51 candidates and screened sera from five patients. In the third stage, the diagnostic potential of 51 candidate peptides was assessed through targeted proteomics of 75 patient sera samples, with a balanced design of 25 malignant pleural mesothelioma, 25 healthy donors, and 25 non-small cell lung cancer patients. Computational analysis found a seven glycopeptide signature for malignant pleural mesothelioma with better discrimination than the FDA approved ELISA assay for mesothelin (Mesomark®).

Targeted proteomics has been applied successfully to formalin-fixed, paraffin-embedded (SFPE) tissue. Sprung and coworkers quantified 114 peptides in FFPE clear cell renal cell carcinomas and Her2 overexpression in FFPE breast cancer samples and determined by comparison with cell lines that lysine-containing peptides can be used for quantitation and the feasibility of performing targeted proteomics studies on FFPE tissues [133]. Takadate and coworkers performed a discovery scan on eight FFPE resectable, node-positive pancreatic ductal adenocarcinoma, and five FFPE noncancerous pancreatic ducts and selected 170 of 1229 candidate proteins for targeted assay development, which they applied to a cohort of 87 cases, finding 14 overexpressed proteins in the poor vs. better outcome groups, ultimately nominating four proteins as prognostic markers: ECH1, OLFM4, STML2, and GTR1 [134]. Pan and coworkers developed assays for five candidate biomarkers of pancreatic cancer and quantified them in plasma obtained from 20 healthy patients, 20 patients with chronic pancreatitis, 20 with early stage pancreatic ductal adenocarcinoma, finding that three of the markers gelsolin, lumican, and tissue inhibitor of metalloproteinase 1 can distinguish pancreatic cancer from controls [135].

5 Conclusion

Targeted proteomics is now a well-established tool for quantitative proteomics and is most useful when researchers can identify a medium-to-large target list (for examples, see Table 1). Through selection of sentinel proteins, the activation state of a given cellular process can be monitored [59], thus extending the coverage of SRM. To date, the technology has been mainly applied to clinical proteomics and less often to basic biological network questions. In the future, targeted proteomics might play a more vital role in characterizing critical post-translationally modified amino acid residues and larger biological networks in a large diversity of developmental stages, disease, contexts, and perturbations. Targeted proteomics should continue to advance in the clinical setting. Targeted proteomics can contribute to an understanding of networks' responses to treatment, such as examination of phosphoprotein response in tumor biopsies before and after treatment. With respect to biomarkers, large collaborative projects such as The Cancer Genome Atlas (TCGA) and CPTAC are finding numerous biomarker candidates, which should be pursued through a sustained commitment to completing the entire biomarker pipeline, which should either eliminate candidates or result in validated and valuable biomarkers translated to clinical practice.

Advances in MS instrumentation will undoubtedly impact the field towards consistently quantifying more proteins per sample. Similar to SRM-MS, PRM-MS (parallel reaction

Table 1. Key studies in targeted MSSome key studies chosen from the reference list covering a wide range of applications of SRM-MS

Study	Description	Assays successfully developed
Biological applications		
Hersmann 2014	Quantitation of cytochrome P450's across developmental stages and tissues	27 cytochrome P450 proteins
Chen 2014	Human liver proteome	57 out of 185 human liver proteins
Worboys 2014	Human kinome	790 proteotypic peptides targeting 196 human kinases – 80% with good quantotypic properties
Wolf-Yadlin 2007	EFGR network across seven time points following EGF stimulation of 184A1 HMEC cells	222 tyrosine phosphopeptides in EGFR network
Sabido 2013	Networks activated by a high fat different across mice strains	144 metabolism related proteins
Bisetto 2013	F_1F_0-ATP synthase super-assembly in H9c2 cardiomyoblasts undergoing differentiation	Complex stoichiometry determined
Kiel 2014	Erbb network in human cancer cell lines	75% of 198 proteins in the network
Clinical applications		
Huttenhain 2013	Glycosites	5568 N-glycosites
Krastins 2013	Samples from seven different clinical areas	16 target proteins spanning pg/ml to ng/ml
Addona 2011	Planned myocardial infarction	121 biomarker candidates
Domanski 2012	Cardiovascular disease	67 candidate biomarkers
He 2014	ERG isoforms in prostate tissue	Multiple ERG isoforms
Whiteaker 2011	Biomarker discovery using a mouse model of breast cancer	88 proteins in 80 plasma samples; 57-plex SRM and 31-plex immuno-SRM
Huttenhain 2012	Cancer-related proteins in plasma and urine	182 proteins in depleted plasma; 408 in urine

monitoring mass spectrometry) takes advantage of a mass spectrometer which first selects precursors using a triple quadrupole and subsequently fragment ions are mass analyzed in an orbitrap [136]. Using the same instrumentation as PRM-MS, the instrument method can be adapted to allow for data independent acquisition [137, 138]. Instead of coupling a quadrupole with an orbitrap mass analyzer, SWATH-MS (sequential window acquisition of all theoretical masses) was implemented on a mass spectrometer which selects the precursors with a quadrupole and all fragment ions are analyzed by TOF [139]. First clinical studies using SWATH-MS demonstrated the usefulness of the method: Liu and colleagues concluded from the analysis of serum proteins of a longitudinal twin study that clinical serum biomarkers should be calibrated against genetic background and age adjusted [140].

We acknowledge Paul Tempst (Memorial Sloan Kettering Center) for thoughtful comments on the manuscript. The work was supported by SystemsX.ch project PhosphoNetX. The project was further supported in part by the SNSF (grant no. 3100A0-688 107679), the European Research Council (grant no. ERC-2008-AdG 233226) to RA.

The authors have declared no conflict of interest.

6 References

[1] Ideker, T., Krogan, N. J., Differential network biology. *Mol. Syst. Biol.* 2012, *8*, 1–9.

[2] Vidal, M., Cusick, M. E., Barabási, A.-L., Interactome networks and human disease. *Cell* 2011, *144*, 986–998.

[3] Bensimon, A., Heck, A. J., Aebersold, R., Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 2012, *81*, 379–405.

[4] Molinelli, E. J., Korkut, A., Wang, W., Miller, M. L. et al., Perturbation biology: inferring signaling networks in cellular systems. *PLoS Comput. Biol.* 2013, *9*, e1003290.

[5] Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I. et al., SAINTexpress: improvements and additional features in significance analysis of interactome software. *J. Proteomics* 2014, *100*, 37–43.

[6] Fermin, D., Basrur, V., Yocum, A. K., Nesvizhskii, A. I., Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* 2011, *11*, 1340–1345.

[7] DeSchutter, E., Why are computational neuroscience and systems biology so separate? *PLoS Comput. Biol.* 2008, *4*, e1000078.

[8] Kiel, C., Ebhardt, H. A., Burnier, J., Portugal, C. et al., Quantification of ErbB network proteins in three cell types using complementary approaches identifies cell-general and cell-type-specific signaling proteins. *J. Proteome Res.* 2014, *13*, 300–313.

[9] Chuang, H.-Y., Hofree, M., Ideker, T., A decade of systems biology. *Annu. Rev. Cell Dev. Biol.* 2010, *26*, 721–744.

[10] Batchelor, E., Loewer, A., Mock, C., Lahav, G., Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* 2011, *7*, 488.

[11] Nilsson, T., Mann, M., Aebersold, R., Yates, J. R., 3rd et al., Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* 2010, *7*, 681–685.

[12] Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A. et al., Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* 2011, *6*, e28766.

[13] Vickaryous, M. K., Hall, B. K., Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Camb. Philos. Soc.* 2006, *81*, 425–455.

[14] Ebhardt, H. A., Sabidó, E., Hüttenhain, R., Collins, B., Aebersold, R., Range of protein detection by selected/multiple reaction monitoring mass spectrometry in an unfractionated human cell culture lysate. *Proteomics* 2012, *12*, 1185–1193.

[15] Omenn, G. S., Plasma proteomics, the Human Proteome Project, and cancer-associated alternative splice variant proteins. *Biochim. Biophys. Acta* 2014, *1844*, 866–873.

[16] Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., Aebersold, R., Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 2009, *138*, 795–806.

[17] Farkash-Amar, S., Eden, E., Cohen, A., Geva-Zatorsky, N. et al., Dynamic proteomics of human protein level and localization across the cell cycle. *PLoS ONE* 2012, *7*, e48722.

[18] Prabakaran, S., Lippens, G., Steen, H., Gunawardena, J., Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2012, *4*, 565–583.

[19] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198–207.

[20] Carlson, S. M., White, F. M., Expanding applications of chemical genetics in signal transduction. *Cell Cycle* 2012, *11*, 1903–1909.

[21] Gajadhar, A. S., White, F. M., System level dynamics of post-translational modifications. *Curr. Opin. Biotechnol.* 2014, *28C*, 83–87.

[22] Hart, Y., Alon, U., The utility of paradoxical components in biological circuits. *Mol. Cell* 2013, *49*, 213–221.

[23] Merbl, Y., Refour, P., Patel, H., Springer, M., Kirschner, M. W., Profiling of ubiquitin-like modifications reveals features of mitotic control. *Cell* 2013, *152*, 1160–1172.

[24] Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K. et al., Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 2006, 441, 840–846.

[25] Purvis, J. E., Lahav, G., Encoding and decoding cellular information through signaling dynamics. *Cell* 2013, *152*, 945–956.

[26] Khersonsky, O., Tawfik, D. S., Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu.Rev. Biochem.* 2010, *79*, 471–505.

[27] Liu, P., Begley, M., Michowski, W., Inuzuka, H. et al., Cell-cycle-regulated activation of Akt kinase by phosphorylation at its carboxyl terminus. *Nature* 2014, *508*, 541–545.

[28] Leng, S. X., McElhaney, J. E., Walston, J. D., Xie, D. et al., ELISA and multiplex technologies for cytokine measurement in inflammation and aging research. *J. Gerontol.* 2008, *63*, 879–884.

[29] Kennedy, J. J., Abbatiello, S. E., Kim, K., Yan, P. et al., Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat. Methods* 2014, *11*, 149–155.

[30] Kuhn, E., Whiteaker, J. R., Mani, D. R., Jackson, A. M. et al., Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Mol. Cell. Proteomics* 2012, *11*, M111.013854.

[31] Wu, Y., Williams, E. G., Dubuis, S., Mottis, A. et al., Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* 2014, *158*, 1415–1430.

[32] Picotti, P., Clement-Ziza, M., Lam, H., Campbell, D. S. et al., A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 2013, *494*, 266–270.

[33] Stergachis, A. B., MacLean, B., Lee, K., Stamatoyannopoulos, J. A., MacCoss, M. J., Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat. Methods* 2011, *8*, 1041–1043.

[34] Gillette, M. A., Carr, S. A., Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* 2013, *10*, 28–34.

[35] Liebler, D. C., Zimmerman, L. J., Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 2013, *52*, 3797–3806.

[36] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* 2012, *9*, 555–566.

[37] Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C. et al., Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* 2014, *13*, 907–917.

[38] Desiderio, D. M., Kai, M., Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed. Mass Spectrom.* 1983, *10*, 471–479.

[39] Ebhardt, H. A., Selected reaction monitoring mass spectrometry: a methodology overview. *Methods Mol. Biol.* 2014, *1072*, 209–222.

[40] Worboys, J. D., Sinclair, J., Yuan, Y., Jorgensen, C., Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nat. Methods* 2014, *11*, 1041–1044.

[41] Barnidge, D. R., Dratz, E. A., Martin, T., Bonilla, L. E. et al., Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.* 2003, *75*, 445–451.

[42] Zhang, F., Bartels, M. J., Stott, W. T., Quantitation of human glutathione S-transferases in complex matrices by liquid chromatography/tandem mass spectrometry with

signature peptides. *Rapid Commun. Mass Spectrom.* 2004, *18*, 491–498.

[43] Heikkinen, A. T., Friedlein, A., Matondo, M., Hatley, O. J. et al., Quantitative ADME Proteomics - CYP and UGT Enzymes in the Beagle Dog Liver and Intestine. *Pharm.Res.* 2015, *32*, 74–90.

[44] Betke, K. M., Rose, K. L., Friedman, D. B., Baucum, A. J., 2nd et al., Differential localization of G protein betagamma subunits. *Biochemistry* 2014, *53*, 2329–2343.

[45] Hersman, E. M., Bumpus, N. N., A targeted proteomics approach for profiling murine cytochrome P450 expression. *J. Pharm. Exp. Ther.* 2014, *349*, 221–228.

[46] Lange, V., Malmstrom, J. A., Didion, J., King, N. L. et al., Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* 2008, *7*, 1489–1500.

[47] Schubert, O. T., Mouritsen, J., Ludwig, C., Rost, H. L. et al., The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. *Cell Host Microbe* 2013, *13*, 602–612.

[48] Chen, C., Liu, X., Zheng, W., Zhang, L. et al., Screening of missing proteins in the human liver proteome by improved MRM-approach-based targeted proteomics. *J. Proteome Res.* 2014, *13*, 1969–1978.

[49] Groh, K. J., Schönenberger, R., Eggen, R. I. L., Segner, H., Suter, M. J.-F., Analysis of protein expression in zebrafish during gonad differentiation by targeted proteomics. *Gen. Comp. Endocrinol.* 2013, *193*, 210–220.

[50] Zulak, K. G., Lippert, D. N., Kuzyk, M. A., Domański, D. et al., Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (Picea abies). *Plant J.* 2009, *60*, 1015–1030.

[51] Choi, S., Kim, J., Yea, K., Suh, P. G. et al., Targeted label-free quantitative analysis of secretory proteins from adipocytes in response to oxidative stress. *Anal. Biochem.* 2010, *401*, 196–202.

[52] Bisson, N., James, D. A., Ivosev, G., Tate, S. A. et al., Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat. Biotechnol.* 2011, *29*, 653–658.

[53] Xiang, Y., Remily-Wood, E. R., Oliveira, V., Yarde, D. et al., Monitoring a nuclear factor-κB signature of drug resistance in multiple myeloma. *Mol. Cell. Proteomics* 2011, *10*, M110.005520.

[54] Sabido, E., Wu, Y., Bautista, L., Porstmann, T. et al., Targeted proteomics reveals strain-specific changes in the mouse insulin and central metabolic pathways after a sustained high-fat diet. *Mol. Syst. Biol.* 2013, *9*, 681-681.

[55] Schmidt, C., Lenz, C., Grote, M., Lührmann, R., Urlaub, H., Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. *Anal. Chem.* 2010, *82*, 2784–2796.

[56] Bisetto, E., Comelli, M., Salzano, A. M., Picotti, P. et al., Proteomic analysis of F1F0-ATP synthase super-assembly in mitochondria of cardiomyoblasts undergoing differentiation to the cardiac lineage. *Biochim. Biophys. Acta* 2013, *1827*, 807–816.

[57] Ori, A., Andres-Pons, A., Beck, M., The use of targeted proteomics to determine the stoichiometry of large macromolecular assemblies. *Methods Cell Biol.*2014, *122*, 117–146.

[58] Ori, A., Banterle, N., Iskar, M., Andres-Pons, A. et al., Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* 2013, *9*, 648.

[59] Soste, M., Hrabakova, R., Wanka, S., Melnik, A. et al., A sentinel protein assay for simultaneously quantifying cellular processes. *Nat. Methods* 2014, *11*, 1045–1048.

[60] Glinski, M., Differential multisite phosphorylation of the trehalose-6-phosphate synthase gene family in *Arabidopsis thaliana*: a mass spectrometry-based process for multi-parallel peptide library phosphorylation analysis. *Mol. Cell. Proteomics* 2005, *4*, 1614–1625.

[61] Chen, V. C., Gouw, J. W., Naus, C. C., Foster, L. J., Connexin multi-site phosphorylation: mass spectrometry-based proteomics fills the gap. *Biochim. Biophys. Acta* 2013, *1828*, 23–34.

[62] Danielson, S. R., Held, J. M., Schilling, B., Oo, M. et al., Preferentially increased nitration of alpha-synuclein at tyrosine-39 in a cellular oxidative model of Parkinson's disease. *Anal. Chem.* 2009, *81*, 7823–7828.

[63] Held, J. M., Danielson, S. R., Behring, J. B., Atsriku, C. et al., Targeted quantitation of site-specific cysteine oxidation in endogenous proteins using a differential alkylation and multiple reaction monitoring mass spectrometry approach. *Mol. Cell. Proteomics* 2010, *9*, 1400–1410.

[64] Huang, F., Zeng, X., Kim, W., Balasubramani, M. et al., Lysine 63-linked polyubiquitination is required for EGF receptor degradation. *Proc. Natl. Acad. Sci.* 2013, *110*, 15722–15727.

[65] Darwanto, A., Curtis, M. P., Schrag, M., Kirsch, W. et al., A modified "cross-talk" between histone H2B Lys-120 ubiquitination and H3 Lys-79 methylation. *J. Biol. Chem.*2010, *285*, 21868–21876.

[66] Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 5860–5865.

[67] Rardin, M. J., Held, J. M., Gibson, B. W., Targeted quantitation of acetylated lysine peptides by selected reaction monitoring mass spectrometry. *Methods Mol. Biol.* 2013, *1077*, 121–131.

[68] Meo, A. D., Diamandis, E. P., Rodriguez, H., Hoofnagle, A. N. et al., What is wrong with clinical proteomics? *Clin. Chem.* 2014, *60*, 1258–1266.

[69] Grebe, S. K., Singh, R. J., LC-MS/MS in the clinical laboratory - where to from here? *Clin. Biochem. Rev.* 2011, *32*, 5–31.

[70] Lawson, A. M., The scope of mass spectrometry in clinical chemistry. *Clin. Chem.* 1975, *21*, 803–824.

[71] Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N., Wang, P., The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* 2008, *2*, 1386–1402.

[72] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 2006, *24*, 971–983.

[73] Ioannidis, J. P. A., Biomarker Failures. *Clin. Chem.*2013, *59*, 202–204.

[74] Abbatiello, S. E., Schilling, B., Mani, D. R., Zimmerman, L. J. et al., Large-scale inter-laboratory study to develop, analytically validate and apply highly multiplexed, quantitative peptide assays to measure cancer-relevant proteins in plasma. *Mol. Cell. Proteomics* 2015, DOI 10.1074/mcp.M114.047050.

[75] Cima, I., Schiess, R., Wild, P., Kaelin, M. et al., Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proc. Natl. Acad. Sci. U S A* 2011, *108*, 3342–3347.

[76] Farrah, T., Deutsch, E. W., Kreisberg, R., Sun, Z. et al., PASSEL: The PeptideAtlas SRMexperiment library. *Proteomics* 2012, *12*, 1170–1175.

[77] Ellis, M. J., Gillette, M., Carr, S. A., Paulovich, A. G. et al., Connecting genomic alterations to cancer biology with proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov.* 2013, *3*, 1108–1112.

[78] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J. et al., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* 2009, *27*, 633–641.

[79] Cox, H. D., Lopes, F., Woldemariam, G. A., Becker, J. O. et al., Interlaboratory agreement of insulin-like growth factor 1 concentrations measured by mass spectrometry. *Clin. Chem.* 2014, *60*, 541–548.

[80] Kuhn, E., Wu, J., Karl, J., Liao, H. et al., Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. *Proteomics* 2004, *4*, 1175–1186.

[81] Anderson, L., Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* 2006, *5*, 573–588.

[82] Keshishian, H., Addona, T., Burgess, M., Kuhn, E., Carr, S. A., Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 2007, *6*, 2212–2229.

[83] Fortin, T., Salvador, A., Charrier, J. P., Lenz, C. et al., Clinical quantitation of prostate-specific antigen biomarker in the low nanogram/milliliter range by conventional bore liquid chromatography-tandem mass spectrometry (multiple reaction monitoring) coupling and correlation with ELISA tests. *Mol. Cell. Proteomics* 2009, *8*, 1006–1015.

[84] Shi, T., Fillmore, T. L., Sun, X., Zhao, R. et al., Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc. Natl. Acad. Sci.* 2012, *109*, 15395–15400.

[85] Shi, T., Gao, Y., Quek, S. I., Fillmore, T. L. et al., A highly sensitive targeted mass spectrometric assay for quantification of AGR2 protein in human urine and serum. *J. Proteome Res.* 2014, *13*, 875–882.

[86] Fallon, J. K., Neubert, H., Hyland, R., Goosen, T. C., Smith, P. C.,Targeted quantitative proteomics for the analysis of 14 UGT1As and -2Bs in human liver using NanoUPLC–MS/MS with selected reaction monitoring. *J. Proteome Res.* 2013, *12*, 4402–4413.

[87] Zhao, Y., Jia, W., Sun, W., Jin, W. et al., Combination of improved (18)O incorporation and multiple reaction monitoring: a universal strategy for absolute quantitative verification of serum candidate biomarkers of liver cancer. *J. Proteome Res.* 2010, *9*, 3319–3327.

[88] Martínez-Morillo, E., Nielsen, H. M., Batruch, I., Drabovich, A. P. et al., Assessment of peptide chemical modifications on the development of an accurate and precise multiplex selected reaction monitoring assay for apolipoprotein e isoforms. *J. Proteome Res.* 2014, *13*, 1077–1087.

[89] Tang, H.-Y., Beer, L. A., Barnhart, K. T., Speicher, D. W., Rapid verification of candidate serological biomarkers using gel-based, label-free multiple reaction monitoring. *J. Proteome Res.* 2011, *10*, 4005–4017.

[90] Tian, Y., Zhou, Y., Elliott, S., Aebersold, R., Zhang, H., Solid-phase extraction of N-linked glycopeptides. *Nat. Protoc.* 2007, *2*, 334–339.

[91] Zhang, H., Li, X. J., Martin, D. B., Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 2003, *21*, 660–666.

[92] Ossola, R., Schiess, R., Picotti, P., Rinner, O. et al., Biomarker validation in blood specimens by selected reaction monitoring mass spectrometry of N-glycosites. *Methods Mol. Biol.* 2011, *728*, 179–194.

[93] Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K. et al., High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell Proteomics* 2007, *6*, 1809–1817.

[94] Zawadzka, A. M., Schilling, B., Held, J. M., Sahu, A. K. et al., Variation and quantification among a target set of phosphopeptides in human plasma by multiple reaction monitoring (MRM) and SWATH MS2 data-independent acquisition. *Electrophoresis* 2014, *35*, 3487–3497.

[95] Huttenhain, R., Surinova, S., Ossola, R., Sun, Z. et al., N-glycoprotein SRMAtlas: a resource of mass spectrometric assays for N-glycosites enabling consistent and multiplexed protein quantification for clinical applications. *Mol. Cell. Proteomics* 2013, *12*, 1005–1016.

[96] Liu, Y., Chen, J., Sethi, A., Li, Q. K. et al., Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acylethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness. *Mol. Cell. Proteomics* 2014, *13*, 1753–1768.

[97] Mause, S. F., Weber, C., Microparticles: protagonists of a novel communication network for intercellular information exchange. *Circ. Res.* 2010, *107*, 1047–1057.

[98] Smalley, D. M., Ley, K., Plasma-derived microparticles for biomarker discovery. *Clin. Lab.* 2008, *54*, 67–79.

[99] Chen, C. L., Lai, Y. F., Tang, P., Chien, K. Y. et al., Comparative and targeted proteomic analyses of urinary microparticles from bladder cancer and hernia patients. *J. Proteome Res.* 2012, *11*, 5611–5629.

[100] Zubiri, I., Posada-Ayala, M., Sanz-Maroto, A., Calvo, E. et al., Diabetic nephropathy induces changes in the proteome of human urinary exosomes as revealed by label-free comparative analysis. *J. Proteomics* 2014, *96*, 92–102.

[101] Kruh-Garcia, N. A., Wolfe, L. M., Chaisson, L. H., Worodria, W. O. et al., Detection of Mycobacterium tuberculosis peptides in the exosomes of patients with active and latent M. tuberculosis infection using MRM-MS. *PLoS One* 2014, *9*, e103811.

[102] Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B. et al., Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). *J. Proteome Res.* 2004, *3*, 235–244.

[103] Dupuis, A., Hennekinne, J.-A., Garin, J., Brun, V., Protein standard absolute quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics* 2008, *8*, 4633–4636.

[104] Oe, T., Ackermann, B. L., Inoue, K., Berna, M. J. et al., Quantitative analysis of amyloidβ peptides in cerebrospinal fluid of Alzheimer's disease patients by immunoaffinity purification and stable isotope dilution liquid chromatography/negative electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2006, *20*, 3723–3735.

[105] Berna, M. J., Zhen, Y., Watson, D. E., Hale, J. E., Ackermann, B. L., strategic use of immunoprecipitation and lc/ms/ms for trace-level protein quantification: myosin light Chain 1, a Biomarker of Cardiac Necrosis. *Anal. Chem.*2007, *79*, 4199–4205.

[106] Nicol, G. R., Han, M., Kim, J., Birse, C. E. et al., Use of an immunoaffinity-mass spectrometry-based approach for the quantification of protein biomarkers from serum samples of lung cancer patients. *Mol. Cell. Proteomics* 2008, *7*, 1974–1982.

[107] Hoofnagle, A. N., Becker, J. O., Wener, M. H., Heinecke, J. W., Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin. Chem.* 2008, *54*, 1796–1804.

[108] Ahn, Y. H., Lee, J. Y., Lee, J. Y., Kim, Y.-S. et al., Quantitative analysis of an aberrant glycoform of TIMP1 from colon cancer serum by L-PHA-enrichment and SISCAPA with MRM mass spectrometry. *J. Proteome Res.* 2009, *8*, 4216–4224.

[109] Neubert, H., Gale, J., Muirhead, D., Online high-flow peptide immunoaffinity enrichment and nanoflow LC-MS/MS: assay development for total salivary pepsin/pepsinogen. *Clin. Chem.* 2010, *56*, 1413–1423.

[110] Neubert, H., Muirhead, D., Kabir, M., Grace, C. et al., Sequential protein and peptide immunoaffinity capture for mass spectrometry-based quantification of total human β-nerve growth factor. *Anal. Chem.* 2013, *85*, 1719–1726.

[111] Kushnir, M. M., Rockwood, A. L., Roberts, W. L., Abraham, D. et al., Measurement of thyroglobulin by liquid chromatography-tandem mass spectrometry in serum and plasma in the presence of antithyroglobulin autoantibodies. *Clin. Chem.* 2013, *59*, 982–990.

[112] Krastins, B., Prakash, A., Sarracino, D. A., Nedelkov, D. et al., Rapid development of sensitive, high-throughput, quantitative and highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum. *Clin. Biochem.* 2013, *46*, 399–410.

[113] Peterman, S., Niederkofler, E. E., Phillips, D. A., Krastins, B. et al., An automated, high-throughput method for targeted quantification of intact insulin and its therapeutic analogs in human serum or plasma coupling mass spectrometric immunoassay with high resolution and accurate mass detection (MSIA-HR/AM). *Proteomics* 2014, *14*, 1445–1156.

[114] Whiteaker, J. R., Zhao, L., Frisch, C., Ylera, F. et al., High-affinity recombinant antibody fragments (Fabs) can be applied in peptide enrichment immuno-MRM assays. *J. Proteome Res.* 2014, *13*, 2187–2196.

[115] Schoenherr, R. M., Saul, R. G., Whiteaker, J. R., Yan, P. et al., Anti-peptide monoclonal antibodies generated for immuno-multiple reaction monitoring-mass spectrometry assays have a high probability of supporting Western blot and ELISA. *Mol. Cell. Proteomics* 2015, *14*, 382–398.

[116] Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J. et al., Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* 2009, *8*, 1860–1877.

[117] Keshishian, H., Addona, T., Burgess, M., Mani, D. R. et al., Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 2009, *8*, 2339–2349.

[118] Addona, T. A., Shi, X., Keshishian, H., Mani, D. R. et al., A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* 2011, *29*, 635–643.

[119] Huillet, C., Adrait, A., Lebert, D., Picard, G. et al., Accurate quantification of cardiovascular biomarkers in serum using protein standard absolute quantification (PSAQ™) and selected reaction monitoring. *Mol. Cell. Proteomics* 2012, *11*, M111.008235.

[120] Domański, D., Percy, A. J., Yang, J., Chambers, A. G. et al., MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* 2012, *12*, 1222–1243.

[121] Yang, S., Chen, L., Sun, S., Shah, P. et al., Glycoproteins identified from heart failure and treatment models. *Proteomics* 2014, *15*, 567–579.

[122] DeSouza, L. V., Taylor, A. M., Li, W., Minkoff, M. S. et al., Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J. Proteome Res.* 2008, *7*, 3525–3534.

[123] Chen, Y., Gruidl, M., Remily-Wood, E., Liu, R. Z. et al., Quantification of β-catenin signaling components in colon cancer cell lines, tissue sections, and microdissected tumor cells using reaction monitoring mass spectrometry. *J. Proteome Res.* 2010, *9*, 4215–4227.

[124] Elschenbroich, S., Ignatchenko, V., Clarke, B., Kalloger, S. E. et al., In-depth proteomics of ovarian cancer ascites: combining shotgun proteomics and selected reaction monitoring mass spectrometry. *J. Proteome Res.* 2011, *10*, 2286–2299.

[125] Remily-Wood, E. R., Liu, R. Z., Xiang, Y., Chen, Y. et al., A database of reaction monitoring mass spectrometry assays for elucidating therapeutic response in cancer. *Proteomics Clin. Appl.* 2011, *5*, 383–396.

[126] Selevsek, N., Matondo, M., Carbayo, M. S., Aebersold, R., Domon, B., Systematic quantification of peptides/proteins in urine using selected reaction monitoring. *Proteomics* 2011, *11*, 1135–1147.

[127] Wang, Q., Chaerkady, R., Wu, J., Hwang, H. J. et al., Mutant proteins as cancer-specific biomarkers. *Proc. Natl. Acad. Sci.* 2011, *108*, 2444–2449.

[128] He, J., Sun, X., Shi, T., Schepmoes, A. A., Fillmore, T. L., Antibody-independent targeted quantification of TMPRSS2-ERG fusion protein products in prostate cancer. *Mol. Oncol.* 2014, *8*, 1169–1180.

[129] Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L. et al., A targeted proteomics–based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* 2011, *29*, 625–634.

[130] Hüttenhain, R., Soste, M., Selevsek, N., Röst, H. et al., Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* 2012, *4*, 142ra194.

[131] Kälin, M., Cima, I., Schiess, R., Fankhauser, N. et al., Novel prognostic markers in the serum of patients with castration-resistant prostate cancer derived from quantitative analysis of the Pten conditional knockout mouse proteome. *Eur. Urol.*2011, *60*, 1235–1243.

[132] Cerciello, F., Choi, M., Nicastri, A., Bausch-Fluck, D. et al., Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. *Clin. Proteomics* 2013, *10*, 16.

[133] Sprung, R. W., Martinez, M. A., Carpenter, K. L., Ham, A.-J. L. et al., Precision of multiple reaction monitoring mass spectrometry analysis of formalin-fixed, paraffin-embedded tissue. *J. Proteome Res.* 2012, *11*, 3498–3505.

[134] Takadate, T., Onogawa, T., Fukuda, T., Motoi, F. et al., Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. *Int. J. Cancer* 2012, *132*, 1368–1382.

[135] Pan, S., Chen, R., Brand, R. E., Hawley, S. et al., Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. *J. Proteome Res.* 2012, *11*, 1937–1948.

[136] Lesur, A., Ancheva, L., Kim, Y. J., Berchem, G. et al., Screening protein isoforms predictive for cancer using immunoaffinity capture and fast LC-MS in PRM mode. *Proteomics Clin. Appl.* 2015, DOI 10.1002/prca.201400158.

[137] Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinovic, S. M. et al., Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen treated 3D liver microtissues. *Mol. Cell. Proteomics* 2015, *14*, 1400–1410.

[138] Gallien, S., Kim, S. Y., Domon, B., Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring. *Mol. Cell. Proteomics* 2015, *14*, 1630–1644.

[139] Gillet, L. C., Navarro, P., Tate, S., Rost, H. et al., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111 016717.

[140] Liu, Y., Buil, A., Collins, B. C., Gillet, L. C. et al., Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* 2015, *11*, 786.

APPENDIX B

APPENDIX - NOTE ON COLLABORATIONS AND AUTHORSHIP

This thesis research was performed in close collaboration with H. Alexander Ebhardt. At the beginning of my thesis research he was a postdoctoral fellow in Ruedi Aebersold's laboratory at the ETH in Zurich. Towards the end of my dissertation research he began his own group at University College Dublin. We held conference calls on a weekly basis and discussed all aspects of my dissertation research. Generally speaking, I performed data analyses and he performed mass spectrometry. We collaborated closely on experimental design and the interpretation of results. Figures 2.1, 2.2, 3.1, and 5.5 were prepared by H. Alexander Ebhardt. MTS assays in figure 5.5A were performed by H. Alexander Ebhardt in his laboratory at University College, Dublin. All other phenotypic assays were performed by me and figures created by me.

Chapter 6 of my thesis research involved collaboration among 3 laboratories. In early 2014 I discussed using AP-MS to identify proteins that interact with androgen receptor with H. Alexander Ebhardt because he had expressed interest in learning the technique in the Aebersold laboratory where there is substantial expertise in AP-MS. I subsequently learned that Wassim Abida in the Sawyers' laboratory was already working on this and we decided to collaborate. The wetlab experiments that comprise chapter 6 were performed by Wassim Abida in the Sawyers laboratory and proteomics by H. Alexander Ebhardt in the Aebersold laboratory. I analyzed the data and created the figures collaboratively with Wassim Abida and H. Alexander Ebhardt. We met and discussed approximately weekly.

APPENDIX C

APPENDIX - MATHEMATICAL MODELING OF DRUG EFFECTS

Mathematical models may be useful abstractions of biological systems that can clarify mechanisms and make predictions. Modeling may be useful for developing drug combinations. By predicting higher order drug combinations from pairs of perturbations an exponentially large search space can be tested first *in silico* before being tested in the laboratory – often erroneously referred to as "validated." *In silico* drug combination screening may be a cost effective way to develop drug combinations.

Modeling perturbation effects has been a central focus of systems biology research and many types of models have been created. Terfve and Saez-Rodriguez presented a four-part taxonmy that group models into "descriptive or predictive", and "network or non-network" [Terfve and Saez-Rodriguez, 2012]. The "descriptive or predictive" dichotomy depends on whether a model is used to make quantitative predictions about the behavior of a system in response to novel perturbations and includes models such as regression, differential equations, and bayesian networks [Terfve and Saez-Rodriguez, 2012]. Hastie and colleagues add an additional distinction: "discriminative models" relate the behavior of some outputs given other inputs, i.e. $P(Y|X)$, or "generative models" that predict the joint behavior of inputs and outputs, i.e. $P(X, Y)$ [Hastie et al., 2009]. Discriminative models have the advantage of being relatively simple. In contrast, generative models have the advantage that they can be sampled from and therefore generate or simulate data. I pursued strategies from several taxa of models for both training and research purposes:

1. predictive non-network: partial-least squares regression

2. descriptive network: partial correlations network

3. predictive network: coupled-system of ordinary differential equations

Each modeling strategy has a set of equations and assumptions; algorithm & numerical implementation; model assessment; predictions & interpretation [Hastie et al., 2009]. I focused predominantly on using pre-existing modeling tools and therefore was concerned with model assessment, predictions and interpretation. Whether a modeling strategy incorporates ”prior knowledge”, such as pathways in databases is another important consideration that might be added to the taxa proposed by Terfve and Saez-Rodrigues. I chose not to incorporate prior knowledge into my modeling, partly for simplicity and partly to compare the structure of networks learned *de novo* from data to databases.

C.1 Partial least-squares regression model

C.1.1 Introduction & Methods

In regression models involving drug responses typically the data are modeled as predicting a phenotype such as cell proliferation or apoptosis $\mathbf{y}$ given an input matrix $\mathbf{X}$ of (phospho)-protein values [Lee et al., 2012]. The general multivariate linear regression model

$$\mathbf{y} = \mathbf{XB} + \epsilon \tag{C.1}$$

requires finding the matrix of coefficients $\mathbf{B}$ that minimizes the error of an objective function. When the objective function is minimizing the sum of squared

error the solution is given by ordinary least-squares

$$\mathbf{B} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \tag{C.2}$$

If the number (phospho)-protein variables exceeds the number of perturbation conditions or if multiple (phospho)-proteins are collinear then $\mathbf{B}$ cannot be found by ordinary least-squares because $\mathbf{X}^\top \mathbf{X}$ is singular or ill-conditioned, i.e., not amenable to downstream analyses [Wehrens and Mevik, 2007]. A way around this issue is to decompose $\mathbf{X}$ and $\mathbf{y}$ into a smaller number of linear combinations of the variables such that as much as possible of the covariance between $\mathbf{X}$ and $\mathbf{y}$ are described by the components of the decomposition [Wehrens and Mevik, 2007, Hastie et al., 2009].

There are several ways to perform the decomposition and obtain partial-least squares regression solutions. Wehrens and Mevik describe a relatively simple iterative method that uses singular-value decomposition (SVD), which is a practically useful type of matrix factorization [Wehrens and Mevik, 2007]. First, the SVD of $\mathbf{S} = \mathbf{X}^\top \mathbf{y}$ is calculated. The first weight vectors consist of the first left and right singular vectors w and q to obtain scores

$$t = \mathbf{X} w \tag{C.3}$$

$$u = \mathbf{Y} q \tag{C.4}$$

The loadings are then obtained by regressing against the $\mathbf{X}$ scores t

$$p = \mathbf{X}^\top t \tag{C.5}$$

$$q = \mathbf{Y}^\top t \tag{C.6}$$

In the next iteration $n+1$ outer products $tp^\top$ and $tq^\top$ are subtracted from the data matrices

$$\mathbf{X}_{n+1} = \mathbf{X}_n - tp^\top \tag{C.7}$$

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n - tp^\top \tag{C.8}$$

The next component is estimated from the SVD of the updated cross-product $\mathbf{X}_{n+1}^\top \mathbf{Y}_{n+1}$ and vectors t,w,u,q and p are saved as columns in matrices **W**, **T**, **Q**, and **P** [Wehrens and Mevik, 2007]. Typically, a user will choose 10 components and use the cross-validation error to choose an optimal number of components. The R package 'pls' provides numerical implementations of several PLSR algorithms [Wehrens and Mevik, 2007]. Partial least-squares regression has been successfully applied to understanding relationships between (phospho)-proteins and drug response [Lee et al., 2012].

C.1.2 Results

The overall quality of the models can be assessed by cross-validation performance on withheld data, e.g., finding the Pearson correlation co-efficient between data and model. In my drug combination studies I used a leave-each-drugs-combinations-out cross-validation scheme used by Molinelli and coworkers [Molinelli et al., 2013]. Because the dataset contains 6 one-drug and 15 two-drug combinations there are 6 pairs of training and test sets used. If model assessment is satisfactory then feature selection can be performed whereby the magnitude of coefficient in the regression model is used to rank order the importance of each (phospho)-protein in predicting drug response phenotypes [Hastie et al., 2009]. Here model performance was poor with Pearson correlation coefficients between data and model averaging 0.22 and the first two components explaining less than 50% of the variance. The exception is for combinations involving temsirolimus with a correlation coefficient of 0.84. Similar results were obtained for integrated dataset with SRM and RPPA mea-

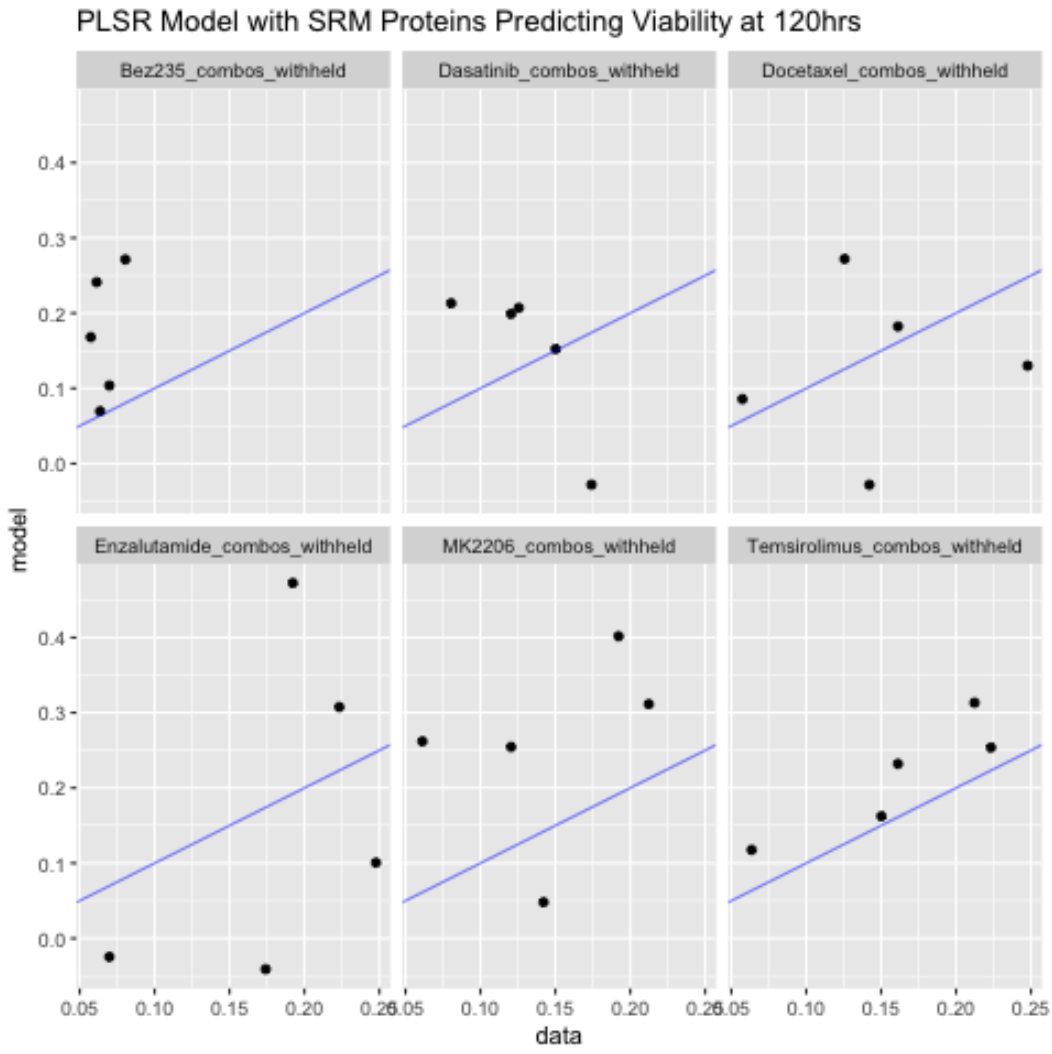

Figure C.1: **PLSR performance on withheld data indicates poor predictive power for most drug combinations except combinations including temsirolimus.**

surements. The models were not deemed to be of sufficient quality to use for feature selection or prediction. There are many other methods for linear regression including random forest, elastic net, lasso, ridge, and principal components regression. I tried a few of them off-the-shelf but did not see a significant increase in performance on withheld data. It is entirely possible that a nonlinear regression method would have better performance.

C.2 Undirected network model with partial correlations

Finding important interactions in data can be challenging due to numerous indirect correlations that can swamp the signal [Stein et al., 2015]. Methods that can clean correlations and identify direct interactions have led to breakthroughs in diverse scientific problems including protein structure prediction through finding coupled amino-acid residues across multiple sequence alignments that form 3D contacts [Marks et al., 2011]. One method to clean correlations is estimation of partial correlations. For random variables a, b, and c rescaled to standard normals the partial correlation between a and b given c is

$$r_{ab|c} = \frac{r_{ab} - r_{bc}r_{ac}}{\sqrt{1 - r_{ac}^2}\sqrt{1 - r_{bc}^2}} \tag{C.9}$$

where r is the Pearson correlation coefficient [Stein et al., 2015]. For multivariate normals the inverse of the empirical covariance matrix gives the matrix of partial correlations. Non-zero partial correlations can be considered edges in a network model, sometimes called gaussian graphical models, concentration graphs, conditional independence graphs, or Markov random fields [Opgen-Rhein and Strimmer, 2007]. Typically there are more (phospho)-protein variables p than measurements n so the empirical covariance matrix cannot be inverted. The GeneNet method uses a sophisticated technique called "shrinkage" that involves shrinking or setting some partial correlations to zero [Opgen-Rhein and Strimmer, 2007, Hastie et al., 2009].

The output from the GeneNet algorithm of edges in the partial correlation network model were filtered for statistical significance with corrected p-values less than 0.05 and the strongest 50 edges were retained for visual clarity. The resulting network model is shown in figure C.2. Some edges are expected based

on prior knowledge. For example, KLK2 and KLK3 are both under androgen regulation; beta-catenin and e-cadherin are important regulations of cell adhesion and motility; pAKT and pIRS1 are connected by an important feedback loop.

Connections in the partial correlations network model may indicate coupling of biological processes and pathway crosstalk. For example, edges among MTOR, cyclinD1 and pPDK1 indicate coordination between growth pathways and cell cycle. Similarly edges among fatty acid synthesis enzymes ACACA, FASN, HSP90 and STAT3 suggest that fatty acid synthesis may be coupled to the stress response and STAT signaling in prostate cancer cells. Co-targeting these pathways with AR or PI3K/AKT inhibitors may enhance their efficacy. A partial correlations network model is a relatively simple network analysis method that is useful for rapid hypothesis generation.

C.3 Coupled-system of nonlinear ordinary differential equations

Network models of biochemical systems that can be used for knowledge discovery and also to predict effects of perturbations are among the most useful [Terfve and Saez-Rodriguez, 2012]. The modeler typically needs to make tradeoffs between model complexity and simplicity. One class of complex models are Michaelis-Menten models. These require detailed knowledge of reactions and fitting of reaction constants and are therefore impractical for new investigations of systems with more than about 20 (phospho)-proteins [Nelander et al., 2008]. A simpler class of models are Hopfield-type networks that were originally

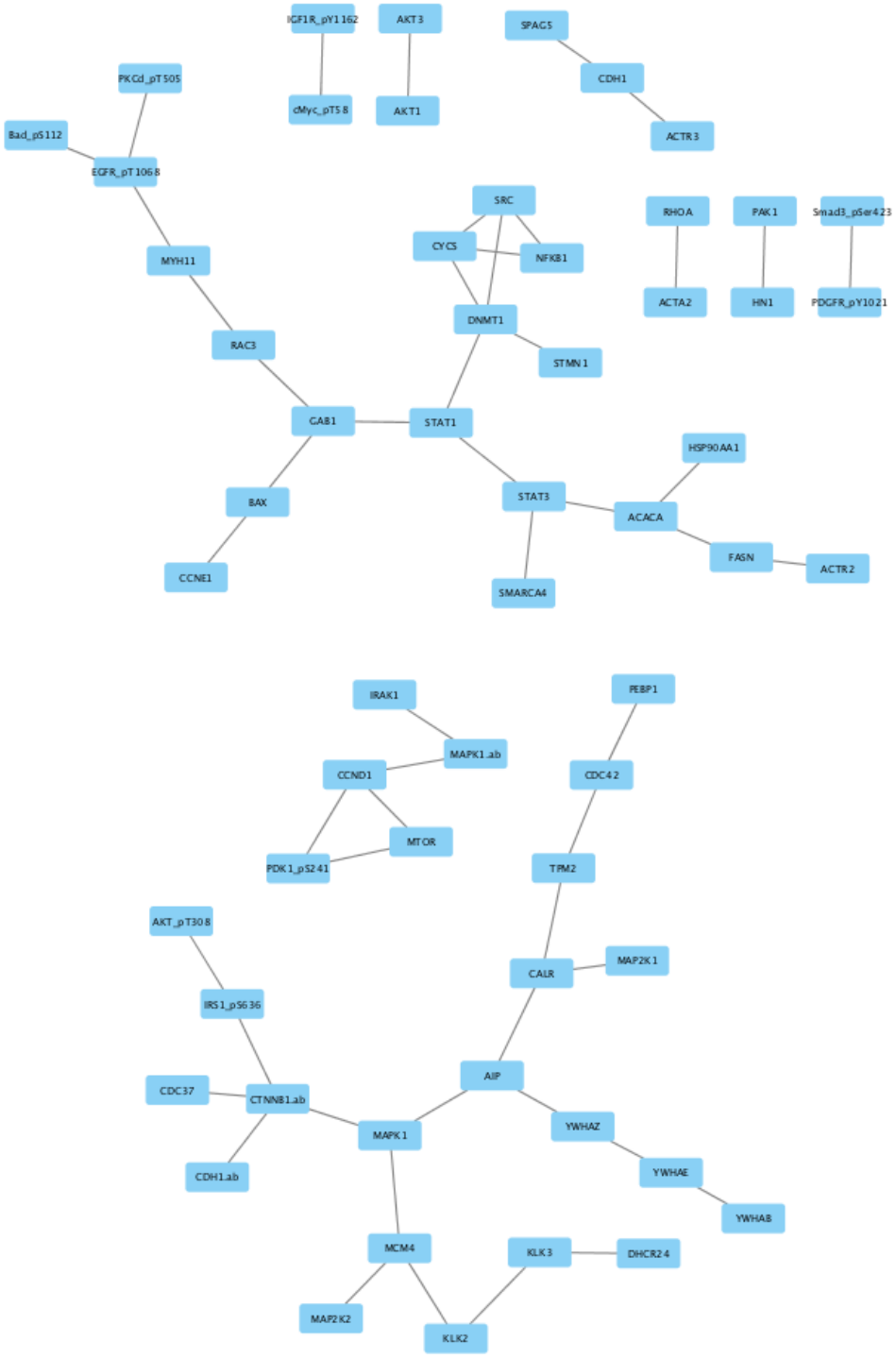

Figure C.2: **Network of partial correlations showing top 50 strongest edges.**

studied in neural networks to represent connectivity, activation, and memory [Hopfield, 1982]. Despite their relative simplicity these networks are capable of complex behaviors present in biological systems, such as non-linearity, feedback, oscillations, memory, robustness, fine-scale tunability, modularity, stable points, stable cycles, bifurcations, apparently random fluctuations, and modeling drug synergy [Hopfield, 1982, Nelander et al., 2008, Molinelli et al., 2013, Miller et al., 2013]. Hopfield-type networks are systems of coupled, non-linear ordinary differential equations. Even very simple systems of first-order finite difference equations can exhibit complex dynamical behavior [May, 1976]. There are challenges in both constructing Hopfield-type network models and analyzing them.

The Hopfield-type network model investigated here was first used in the Sander group by Nelander and colleagues in 2008 and has the form

$$\frac{dx_1^{\mu}}{dt} = \beta_i \arctan\left(\sum_{j\neq i}^{P}(w_{ij}x_j^{\mu}(t) + u_i^{\mu})\right) - \alpha_i x_i^{\mu}(t) \tag{C.10}$$

Where the network structure of p (phospho)-proteins is given by w_{ij} and the time derivative of each (phospho)-protein x_i in perturbation μ is determined by the sum of three terms: the sum of perturbation effect μ on node x_i and its effect on x_i's neighbors x_j, plus a restoring force αx_i that represents the tendency of the system to relax following a perturbation; α and β are constants that modulate the strength of these terms; arctan is a sigmoidal function that models saturation effects.

To fit the models from single timepoint measurements 24 hours following perturbation the system is assumed to be at steady-state. Biologically speaking this may not be correct because several of the drugs are inducing apoptosis. Molinelli and colleagues were able to derive useful and predictive models in

melanoma cells under this assumption [Molinelli et al., 2013]. The perturbation μ can in principal have time dependence but is assumed to be constant here. This means biologically that the intra-cellular drug concentration is not changing over the 24 hour period when the (phospho)-protein measurements were acquired and also not changing over the 72 hour period up to when the models are simululated.

To construct models, an iterative unconstrained global optimization algorithm is required. Models are optimized by minimizing a cost function

$$C(\mathbf{W}) = \beta \sum_{i}^{P} \sum_{\mu}^{N} \left(x_i^{\mu}(t_l) - x_i^{\mu *} \right)^2 + \lambda \sum_{i}^{P} \sum_{j \neq i}^{N} \delta(w_{ij}) \tag{C.11}$$

$$\delta(w_{ij}) = \begin{cases} 1 & \text{if } w_{ij} \neq 0 \\ 0 & \text{if } w_{ij} = 0 \end{cases} \tag{C.12}$$

where the first term represents the sum-of-squares error between model $x_i^{\mu *}$ and data at perturbation μ summed over all P (phospho)-proteins and N perturbations; the second term represents sparsity and an indicating function assigns a 1 to any non-zero interaction term w_{ij}; the relative strength of the error cost is controlled by the constant β and non-sparsity cost by λ, which can be tuned empirically to minimize cost on training data. Construction of these models is considered a challenging problem in machine learning called structure learning [Hastie et al., 2009].

Nelander and colleagues developed their own algorithm COPIA to construct models that uses an iterative combination of Monte-Carlo and gradient calculations called back propagation to minimize the cost function [Nelander et al., 2008]. Back propagation is one of the earliest methods developed to construct Hopfield-type networks or neural networks. The COPIA

method works well in practice but does not run quickly for systems with more than 50 (phospho)-proteins.

Molinelli and colleagues developed a new iterative method for systems with > 50 (phospho)-proteins that constructs Hopfield-type models by implementing a two-part strategy; first, a custom adaptation of the belief propagation algorithm calculates marginal probabilities for each of eleven discrete values for w_{ij}; second, models are instantiated from these marginal probability distributions by initial random sampling followed by a belief propagation-guided decimation algorithm [Molinelli et al., 2013]. The term 'belief propagation' originated in the artificial intelligence research community where 'beliefs' are encoded as probability statements, often of conditional independence between variables, and 'propagation' refers to iteratively updating these probability statements among variables, also termed message passing. Because of the random sampling steps the BP method is run typically 1,000 times to generate an ensemble of models for analysis.

Models constructed by BP can be evaluated by assessing performance on withheld data through a cross-validation scheme. Here I used the same scheme as for partial least-squares regression, i.e., a leave-each-drugs-combinations-out. Because the dataset contains 6 one-drug and 15 two-drug combinations there are 6 pairs of training and test sets used. I used FORTRAN code published by Molinelli and colleagues and wrote additional code in R for data wrangling, visualization, numerically integrating ODE systems to steady-state, and assessing model quality by performance on withheld data. Figure C.3 shows the basic workflow. This workflow was iterated over a grid of sparsity λ and error β values to find the best performance on withheld data.

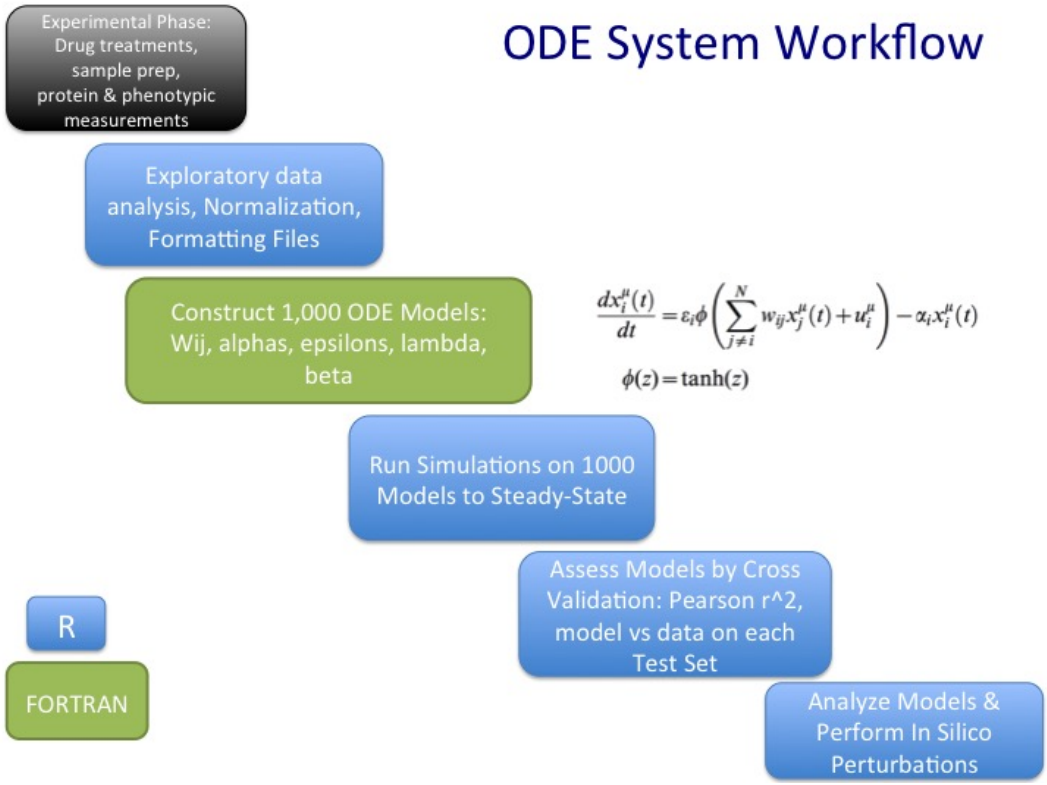

Figure C.3: **Workflow for Hopfield network modeling with belief propagation algorithm, FORTRAN code, and R code.** The flow starts in the upper left corner and moves diagonally to lower right.

Figure C.5 shows the performance of the top ten BP-models with lowest training error on withheld data for the six leave-each-drugs-combos out test sets. Some models achieve an r^2 better than 0.5, however, a minority of models have terrible performance comparable to random guessing. Interestingly, performance in the temsirolimus drug set (6) has far less variance and exceeds a mean of 0.5 r^2. The worst performance is seen in the Bez235 set (1). The same pattern was observed with the performance of the partial least-squares regression models, suggesting that the behavior of some drugs in combination is much easier to predict than others.

The conclusion of model assessment indicated caution in terms of analyzing

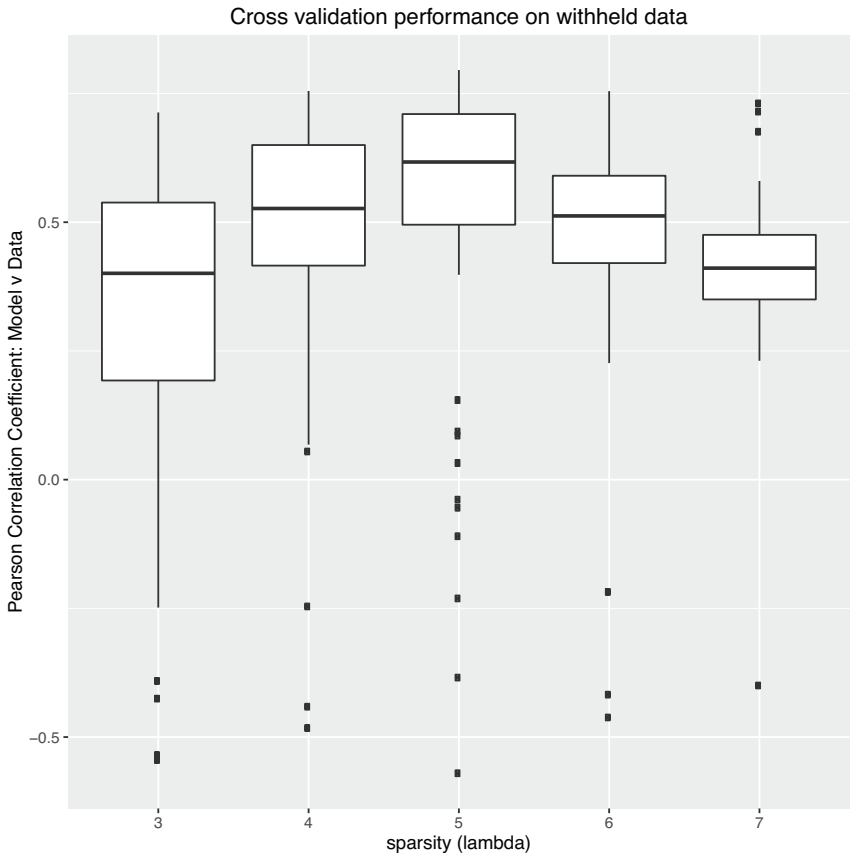

Figure C.4: **Tuning of the sparsity parameter by assessing performance on withheld data for the top 60 models with lowest training error.**

the models and using them to predict novel drug combinations. I ran a number of simulated *in silico* drug perturbations where all pairwise combinations of proteins, targeted by available drugs were explored in terms of their effects on viability and KLK3 (or PSA). The idea here is that PSA often correlates with tumor progression and therefore it could be used as an *in vitro* surrogate of drug efficacy. KLK3 could also be considered a readout of AR transcriptional activity because it is a known target gene of AR. Because AR is considered a driver gene minimizing its activity is also a desirable objective for drug combinations. Figure C.6 shows that there are two-drug perturbation strategies that the ensemble of models predicts will affect KLK3, but no strategies affect cellular phenotypes.

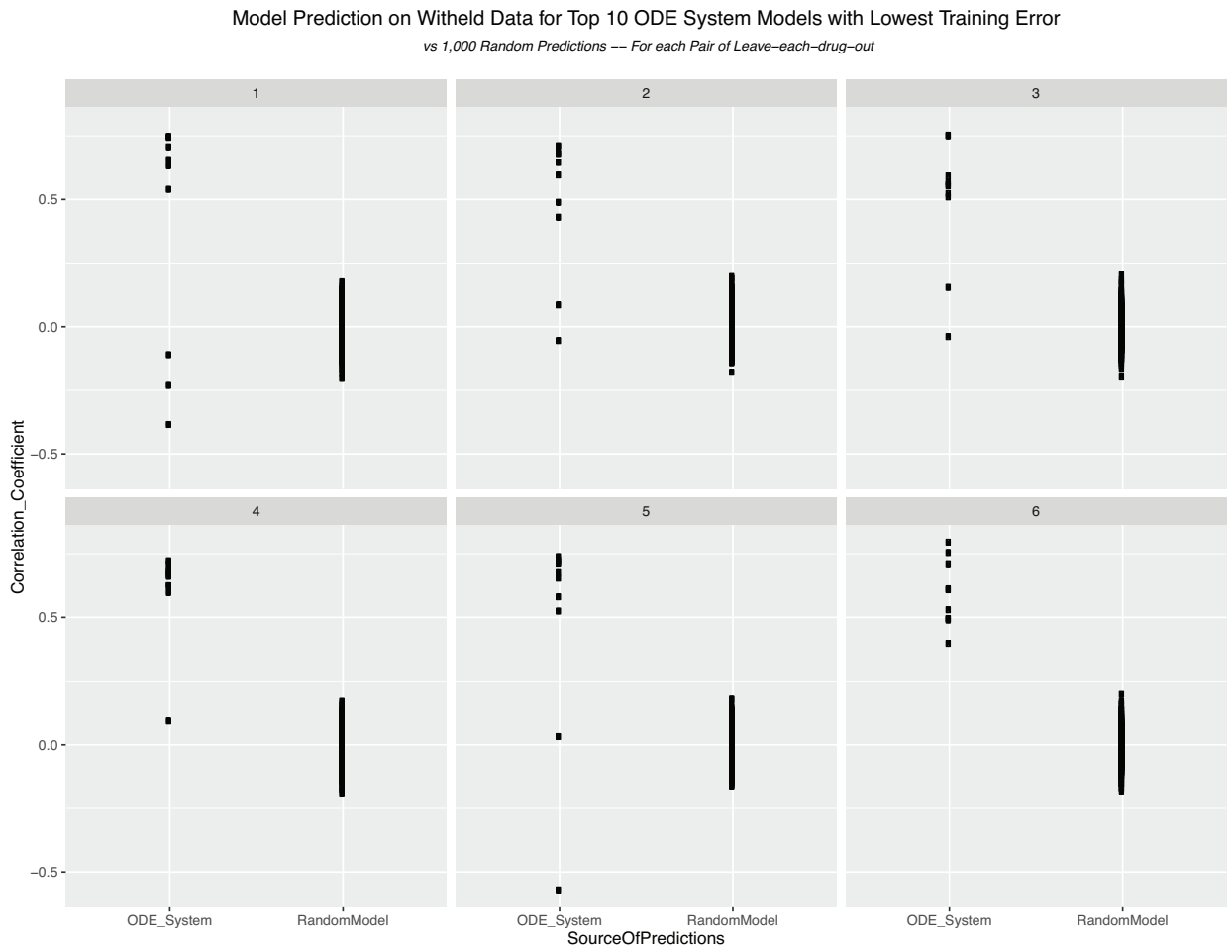

Figure C.5: **Performance of top ten BP-models with lowest training error on withheld data by Pearson correlation data vs model.**

A closer look at two-drug perturbations to minimize KLK3 shown in figure C.7 indicates that the best options include direct inhibition of KLK3 as one of the two drugs in the combination. I evaluated the commercial availability and did not think any of these combinations warranted testing experimentally.

For future directions, it is worth noting that constructing dynamic models from single-timepoint measurements can be error prone [Sachs et al., 2013]. There is often an issue with how accurate the assumption of steady-state behavior is. In addition, Sachs and coworkers showed that if systems

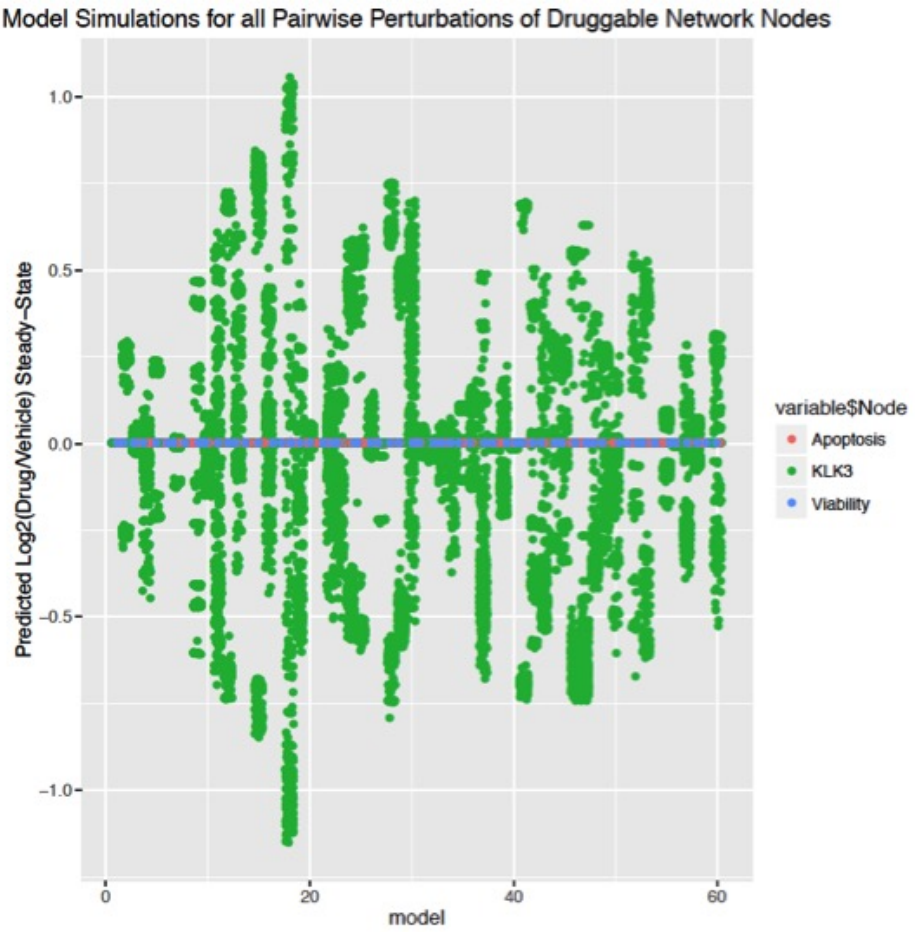

Figure C.6: ***In silico*** **perturbation predictions on cellular phenotypes and KLK3 show no good two-drug combinations to effect cell phenotypes but possibly strategies to minimize KLK3.**

have complicated time-varying behavior then the models that can be learned from single-timepoints may be highly dependent on the choice of timepoint, which is naturally unknown to the experimenters during experimental design [Sachs et al., 2013]. Therefore, there is a necessity in systems biology to develop the experimental will and sample throughput in proteomics to acquire time-series measurements. In his thesis Evan Molinelli worked out extensions of the BP method to construct models based on time-series data [Molinelli, 2013]. This is a promising direction that should be explored further in systems biol-

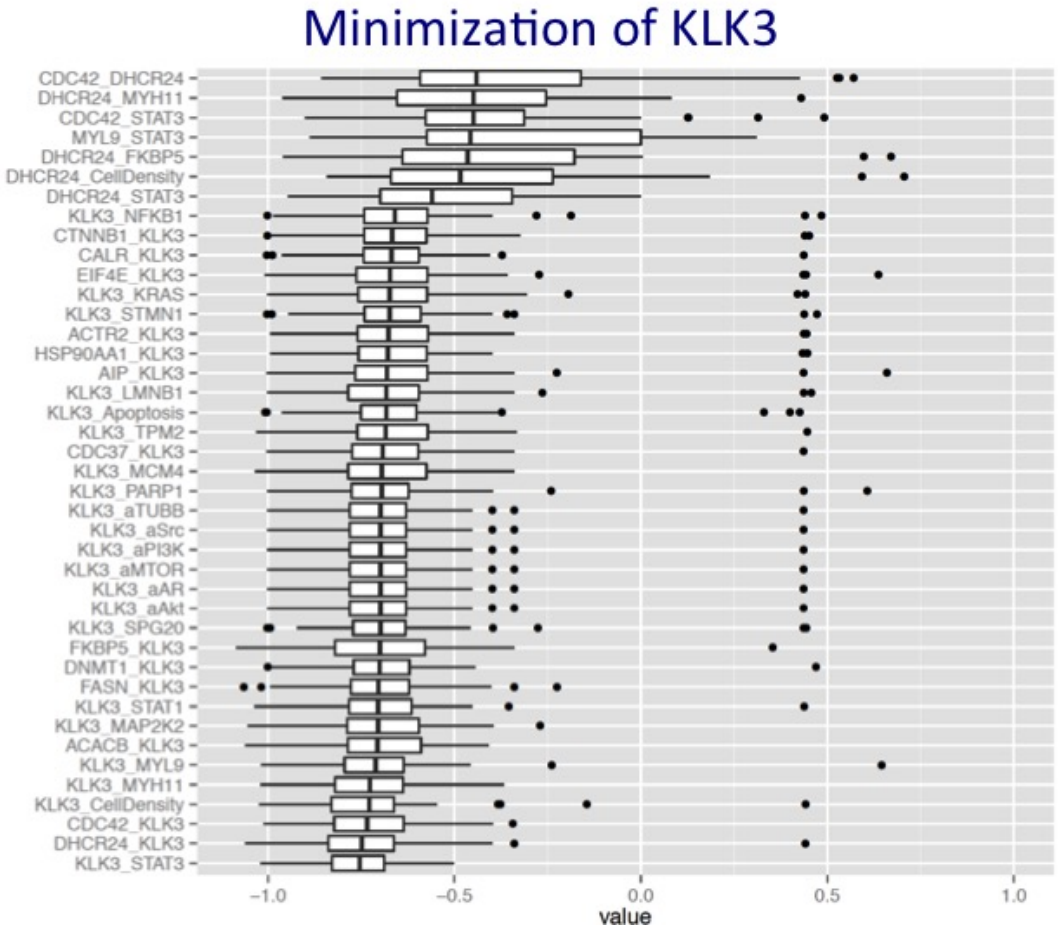

Figure C.7: ***In silico*** **perturbation predictions on KLK3 minimization reveals multiple two-drug combination strategies.**

ogy. The proteomics technology is in place to quantitate (phospho)-proteins either by mass spectrometry or RPPA. Development of PRM assays for cancer-relevant proteins should enable sufficient coverage of important pathways and throughput for systems biology [Cifani and Kentsis,].

BIBLIOGRAPHY

[Aebersold and Mann, 2016] Aebersold, R. and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355.

[Aghazadeh and Papadopoulos, 2016] Aghazadeh, Y. and Papadopoulos, V. (2016). The role of the 14-3-3 protein family in health, disease, and drug development. *Drug discovery today*, 21(2):278–287.

[Aitken, 2006] Aitken, A. (2006). 14-3-3 proteins: A historic overview. *Seminars in Cancer Biology*, 16(3):162–172.

[Anne et al., 2013] Anne, M., Sammartino, D., Barginear, M. F., and Budman, D. (2013). Profile of panobinostat and its potential for treatment in solid tumors: an update. *OncoTargets and therapy*, 6:1613–1624.

[Attard et al., 2016] Attard, G., Parker, C., Eeles, R. A., Schröder, F., Tomlins, S. A., Tannock, I., Drake, C. G., and de Bono, J. S. (2016). Prostate cancer. *Lancet (London, England)*, 387(10013):70–82.

[Au-Yeung et al., 2017] Au-Yeung, G., Lang, F., Azar, W. J., Mitchell, C., Jarman, K. E., Lackovic, K., Aziz, D., Cullinane, C., Pearson, R. B., Mileshkin, L., Rischin, D., Karst, A. M., Drapkin, R., Etemadmoghadam, D., and Bowtell, D. D. L. (2017). Selective Targeting of Cyclin E1-Amplified High-Grade Serous Ovarian Cancer by Cyclin-Dependent Kinase 2 and AKT Inhibition. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23(7):1862–1874.

[Balk, 2007] Balk (2007). AR, the cell cycle, and prostate cancer. *Nuclear Receptor Signaling*, 4.

[Blattner et al., 2017] Blattner, M., Liu, D., Robinson, B. D., Huang, D., Poliakov, A., Gao, D., Nataraj, S., Deonarine, L. D., Augello, M. A., Sailer, V., Ponnala, L., Ittmann, M., Chinnaiyan, A. M., Sboner, A., Chen, Y., Rubin, M. A., and Barbieri, C. E. (2017). SPOP Mutation Drives Prostate Tumorigenesis In Vivo through Coordinate Regulation of PI3K/mTOR and AR Signaling. *Cancer Cell*, 31(3):436–451.

[Bock and Lengauer, 2012] Bock, C. and Lengauer, T. (2012). Managing drug resistance in cancer: lessons from HIV therapy. *Nature reviews. Cancer*, 12(7):494–501.

[Booher et al., 2014] Booher, R. N., Hatch, H., Dolinski, B. M., Nguyen, T., Harmonay, L., Al-Assaad, A.-S., Ayers, M., Nebozhyn, M., Loboda, A., Hirsch, H. A., Zhang, T., Shi, B., Merkel, C. E., Angagaw, M. H., Wang, Y., Long, B. J., Lennon, X. Q., Miselis, N., Pucci, V., Monahan, J. W., Lee, J., Kondic, A. G., Im, E. K., Mauro, D., Blanchard, R., Gilliland, G., Fawell, S. E., Zawel, L., Schuller, A. G., and Strack, P. (2014). MCL1 and BCL-xL levels in solid tumors are predictive of dinaciclib-induced apoptosis. *PLoS ONE*, 9(10):e108371.

[Boutros et al., 2015] Boutros, P. C., Fraser, M., Harding, N. J., de Borja, R., Trudel, D., Lalonde, E., Meng, A., Hennings-Yeomans, P. H., McPherson, A., Sabelnykova, V. Y., Zia, A., Fox, N. S., Livingstone, J., Shiah, Y.-J., Wang, J., Beck, T. A., Have, C. L., Chong, T., Sam, M., Johns, J., Timms, L., Buchner, N., Wong, A., Watson, J. D., Simmons, T. T., P'ng, C., Zafarana, G., Nguyen, F., Luo, X., Chu, K. C., Prokopec, S. D., Sykes, J., Dal Pra, A., Berlin, A., Brown, A., Chan-Seng-Yue, M. A., Yousif, F., Denroche, R. E., Chong, L. C., Chen, G. M., Jung, E., Fung, C., Starmans, M. H. W., Chen, H., Govind, S. K., Hawley, J., D'Costa, A., Pintilie, M., Waggott, D., Hach, F., Lambin, P., Muthuswamy, L. B., Cooper, C., Eeles, R., Neal, D., Tetu, B., Sahinalp, C., Stein, L. D., Fleshner, N., Shah, S. P., Collins, C. C., Hudson, T. J., McPherson, J. D., van der Kwast, T., and Bristow, R. G. (2015). Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nature Genetics*, 47(7):736–745.

[Bozic et al., 2013] Bozic, I., Reiter, J. G., Allen, B., Antal, T., Chatterjee, K., Shah, P., Moon, Y. S., Yaqubie, A., Kelly, N., Le, D. T., Lipson, E. J., Chapman, P. B., Diaz, L. A., Vogelstein, B., and Nowak, M. A. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife*, 2:e00747.

[Bulusu et al., 2016] Bulusu, K. C., Guha, R., Mason, D. J., Lewis, R. P. I., Muratov, E., Kalantar Motamedi, Y., Cokol, M., and Bender, A. (2016). Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug discovery today*, 21(2):225–238.

[Carver et al., 2011] Carver, B. S., Chapinski, C., Wongvipat, J., Hieronymus, H., Chen, Y., Chandarlapaty, S., Arora, V. K., Le, C., Koutcher, J., Scher, H., Scardino, P. T., Rosen, N., and Sawyers, C. L. (2011). Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. *Cancer Cell*, 19(5):575–586.

[Cerami et al., 2012] Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012). The cBio Cancer

Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer discovery*, 2(5):401–404.

[Cerami et al., 2011] Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, Ö., Anwar, N., Schultz, N., Bader, G. D., and Sander, C. (2011). Pathway Commons, a web resource for biological pathway data. *Nucleic acids research*, 39(Database issue):D685–90.

[Chang et al., 2008] Chang, Y.-M., Bai, L., Liu, S., Yang, J. C., Kung, H.-J., and Evans, C. P. (2008). Src family kinase oncogenic potential and pathways in prostate cancer as revealed by AZD0530. *Oncogene*, 27(49):6365–6375.

[Chapman et al., 2016] Chapman, P. J., James, D. I., Watson, A. J., Hopkins, G. V., Waddell, I. D., and Ogilvie, D. J. (2016). IncucyteDRC: An R package for the dose response analysis of live cell imaging data. *F1000Research*, 5:962.

[Chen et al., 2013] Chen, Y., Chi, P., Rockowitz, S., Iaquinta, P. J., Shamu, T., Shukla, S., Gao, D., Sirota, I., Carver, B. S., Wongvipat, J., Scher, H. I., Zheng, D., and Sawyers, C. L. (2013). ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nature medicine*, 19(8):1023–1029.

[Choi et al., 2002] Choi, H., Liu, G., Mellacheruvu, D., Tyers, M., Gingras, A.-C., and Nesvizhskii, A. I. (2002). *Analyzing Protein-Protein Interactions from Affinity Purification-Mass Spectrometry Data with SAINT*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[Choi et al., 2014] Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, 30(17):2524–2526.

[Cifani and Kentsis,] Cifani, P. and Kentsis, A. Towards comprehensive and quantitative proteomics for diagnosis and therapy of human disease.

[Cifani et al., 2017] Cifani, P., Shakiba, M., Chhangawala, S., and Kentsis, A. (2017). ProteoModlR for functional proteomic analysis. *BMC bioinformatics*, 18(1):153.

[Clough et al., 2012] Clough, T., Thaminy, S., Ragg, S., Aebersold, R., and Vitek, O. (2012). Statistical protein quantification and significance analysis in label-

free LC-MS experiments with complex designs. *BMC bioinformatics*, 13 Suppl 16:S6.

[Drake et al., 2016] Drake, J. M., Paull, E. O., Graham, N. A., Lee, J. K., Smith, B. A., Titz, B., Stoyanova, T., Faltermeier, C. M., Uzunangelov, V., Carlin, D. E., Fleming, D. T., Wong, C. K., Newton, Y., Sudha, S., Vashisht, A. A., Huang, J., Wohlschlegel, J. A., Graeber, T. G., Witte, O. N., and Stuart, J. M. (2016). Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell*, 166(4):1041–1054.

[Dumontet and Sikic, 1999] Dumontet, C. and Sikic, B. I. (1999). Mechanisms of Action of and Resistance to Antitubulin Agents: Microtubule Dynamics, Drug Transport, and Cell Death . *Journal of Clinical Oncology*.

[Ebhardt et al., 2015] Ebhardt, H. A., Root, A., Sander, C., and Aebersold, R. (2015). Applications of targeted proteomics in systems biology and translational medicine. *PROTEOMICS*, 15(18):3193–3208.

[Efron and Tibshirani, 2007] Efron, B. and Tibshirani, R. (2007). Accept Terms and Conditions on JSTOR. *The annals of applied statistics*.

[Evans et al., 2016] Evans, J. R., Zhao, S. G., Chang, S. L., Tomlins, S. A., Erho, N., Sboner, A., Schiewer, M. J., Spratt, D. E., Kothari, V., Klein, E. A., Den, R. B., Dicker, A. P., Karnes, R. J., Yu, X., Nguyen, P. L., Rubin, M. A., de Bono, J., Knudsen, K. E., Davicioni, E., and Feng, F. Y. (2016). Patient-Level DNA Damage and Repair Pathway Profiles and Prognosis After Prostatectomy for High-Risk Prostate Cancer. *JAMA oncology*, 2(4):471–480.

[Fong et al., 2009] Fong, L., Kwek, S. S., O'Brien, S., Kavanagh, B., McNeel, D. G., Weinberg, V., Lin, A. M., Rosenberg, J., Ryan, C. J., Rini, B. I., and Small, E. J. (2009). Potentiating Endogenous Antitumor Immunity to Prostate Cancer through Combination Immunotherapy with CTLA4 Blockade and GM-CSF. *Cancer Research*, 69(2):609–615.

[Fu et al., 2000] Fu, H., Subramanian, R. R., and Masters, S. C. (2000). 14-3-3 Proteins: Structure, Function, and Regulation. *Annual Review of Pharmacology and Toxicology*, 40(1):617–647.

[Gao et al., 2016] Gao, S., Gao, Y., He, H. H., Han, D., Han, W., Avery, A., Macoska, J. A., Liu, X., Chen, S., Ma, F., Chen, S., Balk, S. P., and Cai, C. (2016). Androgen Receptor Tumor Suppressor Function Is Mediated by Recruitment of Retinoblastoma Protein. *Cell Reports*, 17(4):966–976.

[Geiger et al., 2012] Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & cellular proteomics : MCP*, 11(3):M111.014050.

[Gottlieb, 1998] Gottlieb, B. (1998). The Androgen Receptor Gene Mutations Database. *Nucleic acids research*, 26(1):234–238.

[Graff et al., 2005] Graff, J. R., McNulty, A. M., Hanna, K. R., Konicek, B. W., Lynch, R. L., Bailey, S. N., Banks, C., Capen, A., Goode, R., Lewis, J. E., Sams, L., Huss, K. L., Campbell, R. M., Iversen, P. W., Neubauer, B. L., Brown, T. J., Musib, L., Geeganage, S., and Thornton, D. (2005). The protein kinase Cbeta-selective inhibitor, Enzastaurin (LY317615.HCl), suppresses signaling through the AKT pathway, induces apoptosis, and suppresses growth of human colon cancer and glioblastoma xenografts. *Cancer Research*, 65(16):7462–7469.

[Gstaiger and Aebersold, 2009] Gstaiger, M. and Aebersold, R. (2009). Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature Reviews Genetics*, 10(9):617–627.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning 2nd edition.

[Hauri et al., 2013] Hauri, S., Wepf, A., van Drogen, A., Varjosalo, M., Tapon, N., Aebersold, R., and Gstaiger, M. (2013). Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. *Molecular Systems Biology*, 9:713.

[Hermeking, 2003] Hermeking, H. (2003). The 14-3-3 cancer connection. *Nature reviews. Cancer*, 3(12):931–943.

[Hopfield, 1982] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558.

[Hsu et al., 2013] Hsu, F.-N., Chen, M.-C., Lin, K.-C., Peng, Y.-T., Li, P.-C., Lin, E., Chiang, M.-C., Hsieh, J.-T., and Lin, H. (2013). Cyclin-dependent kinase 5 modulates STAT3 and androgen receptor activation through phosphorylation of Ser on STAT3 in prostate cancer cells. *American journal of physiology. Endocrinology and metabolism*, 305(8):E975–86.

[Hu et al., 2015] Hu, C., Dadon, T., Chenna, V., Yabuuchi, S., Bannerji, R., Booher, R., Strack, P., Azad, N., Nelkin, B. D., and Maitra, A. (2015). Combined Inhibition of Cyclin-Dependent Kinases (Dinaciclib) and AKT (MK-2206) Blocks Pancreatic Tumor Growth and Metastases in Patient-Derived Xenograft Models. *Molecular cancer therapeutics*, 14(7):1532–1539.

[Huang et al., 2017] Huang, L., Fernandes, H., Zia, H., Tavassoli, P., Rennert, H., Pisapia, D., Imielinski, M., Sboner, A., Rubin, M. A., Kluk, M., and Elemento, O. (2017). The cancer precision medicine knowledge base for structured clinical-grade mutations and interpretations. *Journal of the American Medical Informatics Association : JAMIA*, 24(3):513–519.

[Ioannidis, 2013] Ioannidis, J. P. A. (2013). Biomarker Failures. *Clinical Chemistry*, 59(1):202–204.

[Iorio et al., 2016] Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., van Dyk, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R. K., Liu, Q., Mironenko, T., Mitropoulos, X., Richardson, L., Wang, J., Zhang, T., Moran, S., Sayols, S., Soleimani, M., Tamborero, D., Lopez-Bigas, N., Ross-Macdonald, P., Esteller, M., Gray, N. S., Haber, D. A., Stratton, M. R., Benes, C. H., Wessels, L. F. A., Saez-Rodriguez, J., McDermott, U., and Garnett, M. J. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, 166(3):740–754.

[Ireland et al., 2016] Ireland, L., Santos, A., Ahmed, M. S., Rainer, C., Nielsen, S. R., Quaranta, V., Weyer-Czernilofsky, U., Engle, D. D., Perez-Mancera, P. A., Coupland, S. E., Taktak, A., Bogenrieder, T., Tuveson, D. A., Campbell, F., Schmid, M. C., and Mielgo, A. (2016). Chemoresistance in Pancreatic Cancer Is Driven by Stroma-Derived Insulin-Like Growth Factors. *Cancer Research*, 76(23):6851–6863.

[Jasavala et al., 2007] Jasavala, R., Martinez, H., Thumar, J., Andaya, A., Gingras, A.-C., Eng, J. K., Aebersold, R., Han, D. K., and Wright, M. E. (2007). Identification of putative androgen receptor interaction protein modules: cytoskeleton and endosomes modulate androgen receptor signaling in prostate cancer cells. *Molecular & cellular proteomics : MCP*, 6(2):252–271.

[Komura et al., 2016] Komura, K., Jeong, S. H., Hinohara, K., Qu, F., Wang, X., Hiraki, M., Azuma, H., Lee, G.-S. M., Kantoff, P. W., and Sweeney, C. J. (2016). Resistance to docetaxel in prostate cancer is associated with androgen receptor activation and loss of KDM5D expression. *Proceedings of the National Academy of Sciences*, 113(22):6259–6264.

[Korkut et al., 2015] Korkut, A., Wang, W., Demir, E., Aksoy, B. A., Jing, X., Molinelli, E. J., Babur, Ö., Bemis, D. L., Onur Sumer, S., Solit, D. B., Pratilas, C. A., and Sander, C. (2015). Perturbation biology nominates upstream-downstream drug combinations in RAF inhibitor resistant melanoma cells. *eLife*, 4.

[Kusebauch et al., 2016] Kusebauch, U., Campbell, D. S., Deutsch, E. W., Chu, C. S., Spicer, D. A., Brusniak, M.-Y., Slagel, J., Sun, Z., Stevens, J., Grimes, B., Shteynberg, D., Hoopmann, M. R., Blattmann, P., Ratushny, A. V., Rinner, O., Picotti, P., Carapito, C., Huang, C.-Y., Kapousouz, M., Lam, H., Tran, T., Demir, E., Aitchison, J. D., Sander, C., Hood, L., Aebersold, R., and Moritz, R. L. (2016). Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell*, 166(3):766–778.

[Lapenna and Giordano, 2009] Lapenna, S. and Giordano, A. (2009). Cell cycle kinases as therapeutic targets for cancer. *Nature Reviews Drug Discovery*, 8(7):547–566.

[Lee et al., 2012] Lee, M. J., Albert, S. Y., Gardino, A. K., Heijink, A. M., and Sorger, P. K. (2012). Sequential Application of Anticancer Drugs Enhances Cell Death by Rewiring Apoptotic Signaling Networks. *Cell*.

[Lehár et al., 2009] Lehár, J., Krueger, A. S., Avery, W., Heilbut, A. M., Johansen, L. M., Price, E. R., Rickles, R. J., Short, G. F., Staunton, J. E., Jin, X., Lee, M. S., Zimmermann, G. R., and Borisy, A. A. (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nature biotechnology*, 27(7):659–666.

[Lindqvist et al., 2015] Lindqvist, J., Imanishi, S. Y., Torvaldson, E., Malinen, M., Remes, M., Orn, F., Palvimo, J. J., and Eriksson, J. E. (2015). Cyclin-dependent kinase 5 acts as a critical determinant of AKT-dependent proliferation and regulates differential gene expression by the androgen receptor in prostate cancer cells. *Molecular Biology of the Cell*, 26(11):1971–1984.

[Lu et al., 2016] Lu, Y., Ling, S., Hegde, A. M., Byers, L. A., Coombes, K., Mills, G. B., and Akbani, R. (2016). Using reverse-phase protein arrays as pharmacodynamic assays for functional proteomics, biomarker discovery, and drug development in cancer. *Seminars in oncology*, 43(4):476–483.

[Mancini et al., 2011] Mancini, M., Corradi, V., Petta, S., Barbieri, E., Manetti, F., Botta, M., and Santucci, M. A. (2011). A New Nonpeptidic Inhibitor of 14-3-3 Induces Apoptotic Cell Death in Chronic Myeloid Leukemia Sensitive or

Resistant to Imatinib. *Journal of Pharmacology and Experimental Therapeutics*, 336(3):596–604.

[Marks et al., 2011] Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE*, 6(12):e28766.

[May, 1976] May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*.

[Mellacheruvu et al., 2013] Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin, Z.-Y., Badillo, B. G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A. J. R., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I. M., Bennett, K. L., Washburn, M. P., Raught, B., Ewing, R. M., Gingras, A.-C., and Nesvizhskii, A. I. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature Methods*, 10(8):730–736.

[Menon et al., 2013] Menon, R., Deng, M., Rüenauver, K., Queisser, A., Peifer, M., Pfeifer, M., Offermann, A., Boehm, D., Vogel, W., Scheble, V., Fend, F., Kristiansen, G., Wernert, N., Oberbeckmann, N., Biskup, S., Rubin, M. A., Shaikhibrahim, Z., and Perner, S. (2013). Somatic copy number alterations by whole-exome sequencing implicates YWHAZ and PTK2 in castration-resistant prostate cancer. *The Journal of pathology*, 231(4):505–516.

[Mertins et al., 2016] Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K.-l., Lin, C., McLellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., Carr, S. A., and NCI CPTAC (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, 534(7605):55–62.

[Miller et al., 2013] Miller, M. L., Molinelli, E. J., Nair, J. S., Sheikh, T., Samy, R., Jing, X., He, Q., Korkut, A., Crago, A. M., Singer, S., Schwartz, G. K., and Sander, C. (2013). Drug synergy screen and network modeling in dedifferentiated liposarcoma identifies CDK4 and IGF1R as synergistic drug targets. *Science signaling*, 6(294):ra85.

[Mohammed et al., 2013] Mohammed, H., D'Santos, C., Serandour, A. A., Ali, H. R., Brown, G. D., Atkins, A., Rueda, O. M., Holmes, K. A., Theodorou, V., Robinson, J. L. L., Zwart, W., Saadi, A., Ross-Innes, C. S., Chin, S.-F., Menon, S., Stingl, J., Palmieri, C., Caldas, C., and Carroll, J. S. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Reports*, 3(2):342–349.

[Molinelli, 2013] Molinelli, E. J. (2013). Inferring network models of signal transduction with belief propagation - ProQuest.

[Molinelli et al., 2013] Molinelli, E. J., Korkut, A., Wang, W., Miller, M. L., Gauthier, N. P., Jing, X., Kaushik, P., He, Q., Mills, G., Solit, D. B., Pratilas, C. A., Weigt, M., Braunstein, A., Pagnani, A., Zecchina, R., and Sander, C. (2013). Perturbation biology: inferring signaling networks in cellular systems. *PLOS Comput Biol*, 9(12):e1003290.

[Mooslehner et al., 2012] Mooslehner, K. A., Davies, J. D., and Hughes, I. A. (2012). A cell model for conditional profiling of androgen-receptor-interacting proteins. *International journal of endocrinology*, 2012:381824.

[Nam, 2005] Nam, S. (2005). Action of the Src Family Kinase Inhibitor, Dasatinib (BMS-354825), on Human Prostate Cancer Cells. *Cancer Research*, 65(20):9185–9189.

[Nelander et al., 2008] Nelander, S., Wang, W., Nilsson, B., She, Q.-B., Pratilas, C., Rosen, N., Gennemark, P., and Sander, C. (2008). Models from experiments: combinatorial drug perturbations of cancer cells. *Molecular Systems Biology*, 4(1):216.

[Nesvizhskii, 2014] Nesvizhskii, A. I. (2014). Proteogenomics: concepts, applications and computational strategies : Nature Methods : Nature Research. *Nature Methods*.

[Norris et al., 2009] Norris, J. D., Joseph, J. D., Sherk, A. B., Juzumiene, D., Turnbull, P. S., Rafferty, S. W., Cui, H., Anderson, E., Fan, D., Dye, D. A., Deng, X., Kazmin, D., Chang, C.-Y., Willson, T. M., and McDonnell, D. P. (2009). Differential Presentation of Protein Interaction Surfaces on the Androgen Receptor Defines the Pharmacological Actions of Bound Ligands. *Chemistry & Biology*, 16(4):452–460.

[Oh et al., 2013] Oh, S., Shin, S., Lightfoot, S. A., and Janknecht, R. (2013). 14-3-3 proteins modulate the ETS transcription factor ETV1 in prostate cancer. *Cancer Research*, 73(16):5110–5119.

[Opgen-Rhein and Strimmer, 2007] Opgen-Rhein, R. and Strimmer, K. (2007). From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, 1(1):37.

[Paliouras et al., 2011] Paliouras, M., Zaman, N., Lumbroso, R., Kapogeorgakis, L., Beitel, L. K., Wang, E., and Trifiro, M. (2011). Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes. *Integrative Biology*, 3(10):1020.

[Paruch et al., 2010] Paruch, K., Dwyer, M. P., Alvarez, C., Brown, C., Chan, T.-Y., Doll, R. J., Keertikar, K., Knutson, C., McKittrick, B., Rivera, J., Rossman, R., Tucker, G., Fischmann, T., Hruza, A., Madison, V., Nomeir, A. A., Wang, Y., Kirschmeier, P., Lees, E., Parry, D., Sgambellone, N., Seghezzi, W., Schultz, L., Shanahan, F., Wiswell, D., Xu, X., Zhou, Q., James, R. A., Paradkar, V. M., Park, H., Rokosz, L. R., Stauffer, T. M., and Guzi, T. J. (2010). Discovery of Dinaciclib (SCH 727965): A Potent and Selective Inhibitor of Cyclin-Dependent Kinases. *ACS medicinal chemistry letters*, 1(5):204–208.

[Pauli et al., 2017] Pauli, C., Hopkins, B. D., Prandi, D., Shaw, R., Fedrizzi, T., Sboner, A., Sailer, V., Augello, M., Puca, L., Rosati, R., McNary, T. J., Churakova, Y., Cheung, C., Triscott, J., Pisapia, D., Rao, R., Mosquera, J. M., Robinson, B., Faltas, B. M., Emerling, B. E., Gadi, V. K., Bernard, B., Elemento, O., Beltran, H., Demichelis, F., Kemp, C. J., Grandori, C., Cantley, L. C., and Rubin, M. A. (2017). Personalized In Vitro and In Vivo Cancer Models to Guide Precision Medicine. *Cancer discovery*.

[Polkinghorn et al., 2013] Polkinghorn, W. R., Parker, J. S., Lee, M. X., Kass, E. M., Spratt, D. E., Iaquinta, P. J., Arora, V. K., Yen, W.-F., Cai, L., Zheng, D., Carver, B. S., Chen, Y., Watson, P. A., Shah, N. P., Fujisawa, S., Goglia, A. G., Gopalan, A., Hieronymus, H., Wongvipat, J., Scardino, P. T., Zelefsky, M. J., Jasin, M., Chaudhuri, J., Powell, S. N., and Sawyers, C. L. (2013). Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer discovery*, 3(11):1245–1253.

[Rask-Andersen et al., 2011] Rask-Andersen, M., Almén, M. S., and Schiöth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nature Reviews Drug Discovery*, 10(8):579–590.

[Robinson et al., 2015] Robinson, D., Van Allen, E. M., Wu, Y.-M., Schultz, N., Lonigro, R. J., Mosquera, J. M., Montgomery, B., Taplin, M.-E., Pritchard, C. C., Attard, G., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., Tomlins, S. A., Cooney, K. A.,

Smith, D. C., Brennan, C., Siddiqui, J., Mehra, R., Chen, Y., Rathkopf, D. E., Morris, M. J., Solomon, S. B., Durack, J. C., Reuter, V. E., Gopalan, A., Gao, J., Loda, M., Lis, R. T., Bowden, M., Balk, S. P., Gaviola, G., Sougnez, C., Gupta, M., Yu, E. Y., Mostaghel, E. A., Cheng, H. H., Mulcahy, H., True, L. D., Plymate, S. R., Dvinge, H., Ferraldeschi, R., Flohr, P., Miranda, S., Zafeiriou, Z., Tunariu, N., Mateo, J., Perez-Lopez, R., Demichelis, F., Robinson, B. D., Schiffman, M., Nanus, D. M., Tagawa, S. T., Sigaras, A., Eng, K. W., Elemento, O., Sboner, A., Heath, E. I., Scher, H. I., Pienta, K. J., Kantoff, P., de Bono, J. S., Rubin, M. A., Nelson, P. S., Garraway, L. A., Sawyers, C. L., and Chinnaiyan, A. M. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, 161(5):1215–1228.

[Rodrigues et al., 2017] Rodrigues, D. N., Boysen, G., Sumanasuriya, S., Seed, G., Marzo, A. M. D., and Bono, J. (2017). The molecular underpinnings of prostate cancer: impacts on management and pathology practice. *The Journal of pathology*, 241(2):173–182.

[Rouleau et al., 2017] Rouleau, M., Audet-Delage, Y., Desjardins, S., Rouleau, M., Girard-Bock, C., and Guillemette, C. (2017). Endogenous Protein Interactome of Human UDP-Glucuronosyltransferases Exposed by Untargeted Proteomics. *Frontiers in pharmacology*, 8:23.

[Rüenauver et al., 2014] Rüenauver, K., Menon, R., Svensson, M. A., Carlsson, J., Vogel, W., Andrén, O., Nowak, M., and Perner, S. (2014). Prognostic significance of YWHAZ expression in localized prostate cancer. *Prostate cancer and prostatic diseases*, 17(4):310–314.

[Ruggles et al., 2017] Ruggles, K. V., Krug, K., Wang, X., Clauser, K. R., Wang, J., Payne, S. H., Fenyö, D., Zhang, B., and Mani, D. R. (2017). Methods, Tools and Current Perspectives in Proteogenomics. *Molecular & cellular proteomics : MCP*, 16(6):959–981.

[Ryu et al., 2012] Ryu, B. J., Baek, S.-h., Kim, J., Bae, S. J., Chang, S.-Y., Heo, J.-N., Lee, H., Lee, S. Y., and Kim, S. H. (2012). Anti-androgen receptor activity of apoptotic CK2 inhibitor CX4945 in human prostate cancer LNCap cells. *Bioorganic & medicinal chemistry letters*, 22(17):5470–5474.

[Sachs et al., 2013] Sachs, K., Itani, S., Fitzgerald, J., Schoeberl, B., Nolan, G. P., and Tomlin, C. J. (2013). Single timepoint models of dynamic systems. *Interface focus*, 3(4):20130019.

[Silva et al., 2016] Silva, M. P., Barros-Silva, J. D., Vieira, J., Lisboa, S., Torres, L., Correia, C., Vieira-Coimbra, M., Martins, A. T., Jerónimo, C., Henrique, R.,

Paulo, P., and Teixeira, M. R. (2016). NCOA2 is a candidate target gene of 8q gain associated with clinically aggressive prostate cancer. *Genes, chromosomes & cancer*, 55(4):365–374.

[Siu et al., 2016] Siu, L. L., Lawler, M., Haussler, D., Knoppers, B. M., Lewin, J., Vis, D. J., Liao, R. G., Andre, F., Banks, I., Barrett, J. C., Caldas, C., Camargo, A. A., Fitzgerald, R. C., Mao, M., Mattison, J. E., Pao, W., Sellers, W. R., Sullivan, P., Teh, B. T., Ward, R. L., ZenKlusen, J. C., Sawyers, C. L., and Voest, E. E. (2016). Facilitating a culture of responsible and effective sharing of cancer genome data. *Nature medicine*, 22(5):464–471.

[Skates et al., 2013] Skates, S. J., Gillette, M. A., LaBaer, J., Carr, S. A., Anderson, L., Liebler, D. C., Ransohoff, D., Rifai, N., Kondratovich, M., Težak, Ž., Mansfield, E., Oberg, A. L., Wright, I., Barnes, G., Gail, M., Mesri, M., Kinsinger, C. R., Rodriguez, H., and Boja, E. S. (2013). Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. In *Journal of proteome research*, pages 5383–5394. Biostatistics Center, Massachusetts General Hospital Cancer Center , Boston, Massachusetts 02114, United States.

[Spratt et al., 2016] Spratt, D. E., Zumsteg, Z. S., Feng, F. Y., and Tomlins, S. A. (2016). Translational and clinical implications of the genetic landscape of prostate cancer. *Nature reviews. Clinical oncology*, 13(10):597–610.

[Stein et al., 2015] Stein, R. R., Marks, D. S., and Sander, C. (2015). Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. *PLOS Comput Biol*, 11(7):e1004182.

[Stice et al., 2017] Stice, J. P., Wardell, S. E., Norris, J. D., Yllanes, A. P., Alley, H. M., Haney, V. O., White, H. S., Safi, R., Winter, P. S., Cocce, K. J., Kishton, R. J., Lawrence, S. A., Strum, J. C., and McDonnell, D. P. (2017). CDK4/6 Therapeutic Intervention and Viable Alternative to Taxanes in CRPC. *Molecular cancer research : MCR*.

[Szklarczyk et al., 2015] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(D1):D447–D452.

[Taylor et al., 2010] Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., Reva, B., Antipin, Y., Mitsiades, N., Landers, T., Dolgalev, I., Major, J. E., Wilson, M., Socci, N. D.,

Lash, A. E., Heguy, A., Eastham, J. A., Scher, H. I., Reuter, V. E., Scardino, P. T., Sander, C., Sawyers, C. L., and Gerald, W. L. (2010). Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell*, 18(1):11–22.

[Teo et al., 2014] Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A.-C., and Choi, H. (2014). ScienceDirect. *Journal of Proteomics*, 100(C):37–43.

[Terfve and Saez-Rodriguez, 2012] Terfve, C. and Saez-Rodriguez, J. (2012). Modeling Signaling Networks Using High-throughput Phospho-proteomics. pages 19–57.

[Tran et al., 2009] Tran, C., Ouk, S., Clegg, N. J., Chen, Y., Watson, P. A., Arora, V., Wongvipat, J., Smith-Jones, P. M., Yoo, D., Kwon, A., Wasielewska, T., Welsbie, D., Chen, C. D., Higano, C. S., Beer, T. M., Hung, D. T., Scher, H. I., Jung, M. E., and Sawyers, C. L. (2009). Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science (New York, N.Y.)*, 324(5928):787–790.

[Triantaphyllopoulos et al., 2010] Triantaphyllopoulos, K., Madden, L., Rioja, I., Essex, D., Buckton, J., Malhotra, R., Ray, K., Binks, M., and Paleolog, E. M. (2010). In vitro target validation and in vivo efficacy of p38 MAP kinase inhibition in established chronic collagen-induced arthritis model: a pre-clinical study. *Clinical and experimental rheumatology*, 28(2):176–185.

[Tyanova et al., 2016] Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., and Geiger, T. (2016). Proteomic maps of breast cancer subtypes. *Nature communications*, 7:10259.

[van Bokhoven et al., 2003] van Bokhoven, A., Varella-Garcia, M., Korch, C., Johannes, W. U., Smith, E. E., Miller, H. L., Nordeen, S. K., Miller, G. J., and Lucia, M. S. (2003). Molecular characterization of human prostate carcinoma cell lines. *The Prostate*, 57(3):205–225.

[van Meerloo et al., 2011] van Meerloo, J., Kaspers, G. J. L., and Cloos, J. (2011). Cell Sensitivity Assays: The MTT Assay. In *Cancer cell culture: methods and ...*, pages 237–245. Humana Press, Totowa, NJ.

[Vis et al., 2017] Vis, D. J., Lewin, J., Liao, R. G., Mao, M., Andre, F., Ward, R. L., Calvo, F., Teh, B. T., Camargo, A. A., Knoppers, B. M., Sawyers, C. L., Wessels, L. F. A., Lawler, M., Siu, L. L., Voest, E., and Clinical Working Group of the Global Alliance for Genomics and Health (2017). Towards a global cancer knowledge network: dissecting the current international cancer genomic sequencing landscape. *Annals of Oncology*, 28(5):1145–1151.

[Wang et al., 2011] Wang, H., Zhang, C., Rorick, A., Wu, D., Chiu, M., Thomas-Ahner, J., Chen, Z., Chen, H., Clinton, S. K., Chan, K. K., and Wang, Q. (2011). CCI-779 inhibits cell-cycle G2-M progression and invasion of castration-resistant prostate cancer via attenuation of UBE2C transcription and mRNA stability. *Cancer Research*, 71(14):4866–4876.

[Wang et al., 2017] Wang, J., Ma, Z., Carr, S. A., Mertins, P., Zhang, H., Zhang, Z., Chan, D. W., Ellis, M. J. C., Townsend, R. R., Smith, R. D., McDermott, J. E., Chen, X., Paulovich, A. G., Boja, E. S., Mesri, M., Kinsinger, C. R., Rodriguez, H., Rodland, K. D., Liebler, D. C., and Zhang, B. (2017). Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Molecular & cellular proteomics : MCP*, 16(1):121–134.

[Ward et al., 2015] Ward, R. A., Colclough, N., Challinor, M., Debreczeni, J. E., Eckersley, K., Fairley, G., Feron, L., Flemington, V., Graham, M. A., Greenwood, R., Hopcroft, P., Howard, T. D., James, M., Jones, C. D., Jones, C. R., Renshaw, J., Roberts, K., Snow, L., Tonge, M., and Yeung, K. (2015). Structure-Guided Design of Highly Selective and Potent Covalent Inhibitors of ERK1/2. *Journal of medicinal chemistry*, 58(11):4790–4801.

[Watson et al., 2015] Watson, P. A., Arora, V. K., and Sawyers, C. L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nature reviews. Cancer*, 15(12):701–711.

[Wehrens and Mevik, 2007] Wehrens, R. and Mevik, B. H. (2007). The pls package: principal component and partial least squares regression in R.

[Wei et al., 2015] Wei, Y.-B., Guo, Q., Gao, Y.-L., Yan, B., Wang, Z., Yang, J.-R., and Liu, W. (2015). Repression of metadherin inhibits biological behavior of prostate cancer cells and enhances their sensitivity to cisplatin. *Molecular medicine reports*, 12(1):226–232.

[Yap et al., 2011] Yap, T. A., Yan, L., Patnaik, A., Fearen, I., Olmos, D., Papadopoulos, K., Baird, R. D., Delgado, L., Taylor, A., Lupinacci, L., Riisnaes, R., Pope, L. L., Heaton, S. P., Thomas, G., Garrett, M. D., Sullivan, D. M., de Bono, J. S., and Tolcher, A. W. (2011). First-in-Man Clinical Trial of the Oral Pan-AKT Inhibitor MK-2206 in Patients With Advanced Solid Tumors. *Journal of Clinical Oncology*, 29(35):4688–4695.

[Zaman et al., 2014] Zaman, N., Giannopoulos, P. N., Chowdhury, S., Bonneil, E., Thibault, P., Wang, E., Trifiro, M., and Paliouras, M. (2014). Proteomic-coupled-network analysis of T877A-androgen receptor interactomes can pre-

dict clinical prostate cancer outcomes between White (non-Hispanic) and African-American groups. *PLoS ONE*, 9(11):e113190–e113190.

[Zekri et al., 2015] Zekri, A., Ghaffari, S. H., Ghanizadeh-Vesali, S., Yaghmaie, M., Salmaninejad, A., Alimoghaddam, K., Modarressi, M. H., and Ghavamzadeh, A. (2015). AZD1152-HQPA induces growth arrest and apoptosis in androgen-dependent prostate cancer cell line (LNCaP) via producing aneugenic micronuclei and polyploidy. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 36(2):623–632.

[Zhang et al., 2016] Zhang, Y., Jin, Z., Zhou, H., Ou, X., Xu, Y., Li, H., Liu, C., and Li, B. (2016). Suppression of prostate cancer progression by cancer cell stemness inhibitor napabucasin. *Cancer medicine*, 5(6):1251–1258.

[Zhao et al., 2011] Zhao, J., Meyerkord, C. L., Du, Y., Khuri, F. R., and Fu, H. (2011). 14-3-3 proteins as potential therapeutic targets. *Seminars in Cell & Developmental Biology*, 22(7):705–712.

[Zimmer et al., 2016] Zimmer, A., Katzir, I., Dekel, E., Mayo, A. E., and Alon, U. (2016). Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proceedings of the National Academy of Sciences*, 113(37):10442–10447.